Naval Ocean Systems Center
San Diego, CA 92152-5000

(2)

Technical Document 2216
November 1991

Modeling Choice Under Uncertainty in Military Systems Analysis

P. E. Girard
Sciences Applications International Corp.

M. F. O'Connor
J. W. Ulvila
Decision Science Consortium, Inc.

NAVAL OCEAN SYSTEMS CENTER
San Diego, California 92152-5000

J. D. FONTANA, CAPT, USN
Commander

R. T. SHEARER, Acting
Technical Director

ADMINISTRATIVE INFORMATION

The work reported here was conducted for Chief of Naval Operations, Deputy Chief for Naval Warfare, during FY 91 as part of the Naval Warfare Analysis Program. This document presents the results of a collaborative effort involving Science Applications International Corporation; Decision Science Consortium, Inc.; and NOSC Systems Analysis Group (Code 171) personnel. The contractor work was conducted under contract number N66001-90-D-0192.

Released by
G. L. Ruptier, Project Manager

Under authority of
J. T. Avery, Head
Systems Analysis Group

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

1.0 INTRODUCTION

Within the organization of the Deputy Chief of Naval Operations for Naval Warfare, the Director, Space and Electronic Warfare (OP-76) is responsible for the Space and Electronic Warfare (SEW) Appraisal, a major component of the Navy's Planning, Programming, and Budgeting System. OP-76 is also responsible for the administration of Team "C", which is guiding the development of the Navy's SEW Master Plan. This plan includes Command, Control and Communications requirements. OP-76 is evolving a methodology for analyzing SEW requirements in support of these efforts.

The principal philosophy of the program has been to try to codify the elements of assessment, to understand the relationship of performance to worth in terms of objectives, and to identify the contribution of decision making support to operational outcomes. The results are that a few simple ideas can be used to analyze a very complex system:

(1) The decision to be addressed is how to choose among alternative system configurations, consisting of numbers of forces and their capabilities.

(2) Top Level Warfare Requirements are the basis for judgment, but need to be stated not only by Mission Success Criteria but also in terms of the worth of those outcomes.

(3) Situation settings (scenarios, force laydowns, etc.) establish initial and boundary conditions.

(4) Capabilities are expressed as probabilities of low-level outcomes conditioned on system configuration, the environment, and the situation, including enemy parameters.

(5) All functions, including procedures and tactics, and their performance can be represented by conditional logic models (or modules), which can be chained together, in a mutually supportive hierarchy of objectives and functions, to model the overall process as a causal conditional relationship structure.

(6) Operational decision functions are conditioned on information inputs from physical systems and they control other decisions, as well as physical systems, and are necessary for the activation of other functions. They can be modeled in the same way as other functions, i.e, as conditional objects.

(7) The model should produce distributions of the TLWR outcomes. These and other criteria, such as cost, are then weighted by their worth. The net worth for the alternatives is compared to judge the best alternative.

The problem must be bounded by some top-level objectives, even though there are always higher level objectives being served. In the case of Naval Warfare, these objectives have been stated in terms of the Top-Level Warfare Requirements (TLWR). These are stated in terms of enemy and own force losses under hostile situations or in terms of information known about the enemy in peacetime or crisis conditions. Systems are described in terms of their performance of certain characteristic functions which that type of system performs. The problem of assessing system effectiveness is in being able to relate the systems' performance to the TLWR. The problem of assessing a system's worth in comparison to another is in determining the difference that their performance would make in terms of the worth assigned to the Mission Success Criteria of the TLWR.

Today, we have neither the tools nor the information necessary to make either an effectiveness or a worth judgment. First, there is no model that can predict the outcome of the situations described in the TLWR. Second, no worth judgment has been incorporated in the TLWR process on which to base comparisons, e.g., how much more is the attrition of n+1 submarines worth than that of n submarines?

In the case of modeling, we do not have a well organized discipline of system description and characterization. First of all, we do not express performance in terms of what a system actually does. For example, a sonar does not have an intrinsic range; it has a threshold signal-to-noise ratio. Yet we state its performance in terms of range, which also depends on the characteristics of the enemy and the environment. In addition, we only know how to deal with the simplest of functions in terms of performance. In particular, little more than probability of detection and probability of kill is dealt with, effectively, in models. Counter-detection and counter-kill measures are also treated. Maneuver is treated deterministically, which is not too poor an assumption. But unfortunately, decisions are also preprogrammed in models as deterministic activation of the other functions. This prohibits any tradeoff in the decision support systems to effect a different result in terms of TLWRs. In order to assess C3 systems, we must be able to model how likely the decision maker is to recognize the situation, how likely that person is to choose particular courses of action as a result of that recognition and when these events will take place. Finally, the complexity of the problem makes relating performance measures at the system level to the outcomes at the TLWR level a monumental task; one which requires a breakthrough in model organization and computation. Lacking this overall modeling capability, we resort to simpler models which cannot resolve the differences in outcomes caused by C3 systems. Another fallback position is to state some system capabilities such as "throughput" and "capacity" or ill-defined terms, such as "timeliness", as substitutes for the TLWR objectives. This would be valid if "more is better" were directly related to better performance of TLWRs. But this is not necessarily the case. Witness, in Desert Shield/Storm, more throughput inundated command centers with redundant reports and irrelevant information for that command.

Even if we could express the capabilities of C3 at the system level, as we do detection and kill, it is noted, in Section 2.3, that worth cannot be ascertained and applied at that level, unless it can be demonstrated that the worth at that level relates to the worth at the TLWR level. The basis for this relation is shown to be the model of the outcomes given the capabilities. Section 2.4 explains that decision trees and influence diagrams are representations of the conditional probability of a chain of events and are, in effect, a picture of the model relation between capabilities and outcomes. The ability to model a problem as complex as the set of situations forming the basis of the TLWRs is indeed a challenge. New initiatives are needed in this area.

Even if we had the capability to model the effects of performance on the achievement of TLWR outcomes, the evaluation of systems also requires a comparison of the relative worth of those outcomes. Since the outcomes are stated in terms of multiple objectives, e.g., our losses vs. their losses of various types of platforms or personnel, the relative merit of different combinations of the objective outcomes must be known. This means that TLWRs need to be augmented with the tradeoff values of these objectives. These should be determined during the TLWR definition process.

But still, the inability to model the effects leaves a significant gap in the process. The questions to be addressed are (1) what characteristic functions and performance measures of a C3 system can be used to describe how well the decision process is supported by that system, (2) what is the perceived value of combinations of these performance measures, and (3) how does this value relate to the value of the likely outcomes in terms of TLWRs?

These questions will have to be left to later work to fill in the details. This report presents the rationale for a method of analysis and assessment, based on Multi-Attribute Utility Theory

(MAUT), which identifies the elements of the problem and places into perspective the aspects of objectives, system capabilities, choice alternatives and their relationships. In addition, several other methods will be addressed, which do not provide a sound foundation or basis for their usefulness, primarily since they do not have an optimization criterion on which to base judgments, as MAUT does. But these methods get mentioned as approaches that avoid the problems of complexity of MAUT or that they are "intuitively appealing". Intuition is not a basis for judging a method and avoiding complexity does not solve the problem. However, when the problem has not been solved, some approximations and uncertainty remain. Then the method chosen must provide insight into the degree of completeness and reliance one can reasonably place in the results. Some of these methods attempt to approximate the results, but a good measure of validity is lacking.

The criterion for judgment in MAUT is the Expected Utility of certain outcomes, which will be defined in Section 2.1, for single attributes, and extended for multiple attributes in Section 2.2. This criterion provides a common basis of comparison among choice alternatives. It does not, however, provide a measure of robustness in that there is no indication of the sensitivity or dispersion of the result. The decision maker wants to have some idea how variable the outcome might be. A result with a higher expected value is not necessarily better than a result with lower expected value, but a tighter bound on the outcome. The MAUT method does, however, have the necessary information to provide a measure of dispersion, but in practice it has not been recognized because decision analysts have not considered the results in the form of the probability distribution of the value (worth or utility) of the outcome. Section 2.5 defines a method for producing this result.

In previous work (NOSC TD 1938, Girard 1989A, 1990), a method of modeling has been developed, which applies to all levels of complexity and incorporates decision making as an enabling function. This modeling approach recognizes a Hierarchy of Objectives, originating with Mission Success Criteria and expanding to lower level objectives in a Mission, Functions, and Tasks support structure, ending with system capabilities. The relationship connecting the elements of the model is causal conditioning and the operation of the model is based on Conditional Probability Logic.

<u>When the Hierarchy of Objectives Model, implemented with Conditional Probability Logic, is combined with Multi-Attribute Utility Theory, a Model of Choice Under Uncertainty is the result.</u> The performance (and cost) models are inserted between the set of alternatives of the choice, including their capabilities, and the top level objectives, with utility functions applied to the important attributes associated with those objectives. This report describes how to integrate these approaches.

1.1 ORGANIZATION OF THE REPORT

The report consists of an executive overview of the MAUT approach, expanded with conditional modeling, and a summary of other techniques and concludes with an example that illustrates the elements of the process and a set of recommendations for conducting system assessment. There is also an appendix that goes into the details of the techniques which were reviewed. A second appendix lists some research issues highlighted in the course of the effort. References for the entire document are provided at the end of the main body.

2.0 OVERVIEW OF MULTI-ATTRIBUTE UTILITY THEORY

MAUT is the only normative model of rational choice using probability and utility to find expected worth as a basis to compare alternatives. As a rational model, it identifies the best of the alternatives in the sense that if this were a betting situation, it identifies the best bet. Any other choice must be expected to pay off with less value on the average. But it has been recognized that MAUT is not a descriptive theory of how people actually make decisions. For this reason, there are other models being investigated, which may be more descriptive of what people do. These may eventually represent people well, but they would not be good normative models. The situation is analogous to gambling. If the choice is whether to gamble or not, then clearly, when the odds are in favor of the house, the rational answer would be not to gamble. But people do gamble, partially because of the fun, which could be accounted for in a utility sense, but also because they think about the possibility that they may beat the odds. This, too, can be accounted for, in subjective expected utility, by a model which says that the decision maker believes that for them the odds will be shifted in their favor, even when, rationally, they know better. So people do not necessarily choose on the basis of the expected value, but on that of the desired value. This is a case where the desire for an outcome shifts the subjective probabilities of its occurrence. Another reason people may not act as the expected utility model recommends is that the kind of decisions to be made will be made once and only once, so there are not a large number of trials over which to average the outcome. Decision makers can not expect the average result when there is only one opportunity. Nonetheless, the alternative with the best expected value is still the rational choice, because then the outcomes of less value are balanced by the better ones. The question is whether people can or should be convinced to make the normative decision rather than the intuitive or biased or expedient one.

Another important reason for people not making the decision suggested by MAUT is that they may not have a good model of the likelihood of the outcomes that would accrue for each choice that is available. And, due to bounded rationality, complex problems may exceed human capacity to analyze the problem without mechanical help from computer models. The issue of the role of models will be seen to be very important to the decision process. While MAUT seems to have been substantiated for small problems, its extension to complex problems requires significant modeling capability. This section will review the elements of MAUT.

Before proceeding, some terms need to be discussed, which are ambiguous and confusing, in that they have several meanings that are correct in different senses when applied to decision analysis.

VARIABLE: The set of variables in the problem or state space includes all physical descriptors, operational decisions, utilities and situation descriptors. The set of choices among which to choose is a variable, too. The term "variable" includes the special case when its value is constant.

OUTCOME: Any final (or intermediate) instance of the combination of values of the variables in the problem may be called the outcome (or intermediate result or outcome).

STATE: The word "state" may refer to a variable name or it may mean the instance (value) of the variable. It may also refer to a specific outcome or combination of the variables of a system.

VALUE/WORTH/UTILITY: The word "value" has two meanings that confuse the issue. It may mean the worth in monetary or other absolute terms, or it may mean the instance or level of a variable. Utility is related to the monetary or absolute value, but is scaled to lie in the range of zero to one for the range of the outcomes that might occur, i.e., the highest valued outcome is given a utility of one and the least valued gets utility zero. All other outcomes get intermediate utility scores proportional to their value between the highest and lowest. "Value", "worth" and "utility" all refer to payoff ideas. Note that worth is a variable which may be instantiated at a particular value (of worth) and similarly for utility. (See Keeney and Raiffa, 1976, Chapters 3 and 6, for a discussion of value and utility functions.) "Value", "instance" or "level" will refer to the outcome of variables or states.

ATTRIBUTE: This word is used in three different ways in the systems analysis and decision theory areas. Object oriented (and relational) programming uses the term "attribute" to mean any (significant, global or key) variable. Some people use it to mean a characteristic or quality, without being specific about the nature of the quality. Decision theorists use it to mean a variable which is an objective (one which has varying utility) in the decision. Another term with this meaning is criterion. In a context free sense, any variable can be an attribute, in particular for programmers and modelers. But some variables, like decision states, intermediate outcomes, and utilities, do not have worth attached and are not candidates for the decision theoretic term, attribute. Utility or worth is attached to the important attributes.

In Decision Analysis, the "attributes" are only those (state) variables that have worth in the context of the decision (choice among alternatives). In terms of Naval Warfare, the Mission Succes Criteria in the Top Level Warfare Requirements are the only attributes. In a higher level of objectives, peace and economic factors may be the criteria of interest.

2.1 UTILITY THEORY

Utility Theory is used to describe the motivation for decision making. Given a set of choices (alternative actions or objects to select), certain outcomes will result, depending on the decision to choose one or another of the alternatives. These outcomes are the set of possible values of attributes or states of the "system" or "situation" with which the decision maker is concerned. Utility Theory applies a measure of "worth" to each value of a single attribute. For each choice of alternative course of action, there is a probability distribution believed to hold with regard to which value of the attribute will be achieved. The expected "utility" of choosing that alternative is the sum of the products of the "worth" of an outcome value times the probability of that outcome, given that that alternative is chosen. The alternative with the highest expected utility should be the "best" choice. That is the fundamental principle upon which Utility Theory is based, that the decision maker <u>should</u> choose the alternative which maximizes expected utility.

Suppose there is a set of alternatives, A, an event (or quality or state) attribute, X, and that there is a utility function concerning the attribute, u(x), and a probability of occurrence of particular outcomes of the attribute, $p(x \mid A)$, that depends on the choice of the alternative. The expected utility of choosing each alternative is

$$EU(A) = \int u(x) \cdot p(x \mid A)\, dx \tag{1}$$

Note that the utility function does not depend on the alternatives, only on the outcomes of the attribute, whereas the probability of the outcomes is conditional on the alternative chosen. Equation (1) represents the basic postulate of Decision Analysis, that people seek to optimize the objective function of expected utility. Sometimes the objective function is expressed as Subjective Expected Utility. There are two aspects to the meaning of "subjective". One is that the value function or the utility function is subjective. The probability function may also be subjective, if there is no completely empirical evidence for the model of the outcome given the choice. An important issue in eliciting peoples' judgments about the utility and probability functions is whether they separate the ideas in their responses. They may bias their probability judgment toward the desirable region of the attribute domain. They may also bias their probability judgment in favor of the alternative they desire for reasons other than the attainment of the stated attribute levels.

2.2 MULTI-ATTRIBUTE UTILITY THEORY

Since outcomes do not always involve single attributes, Multi-Attribute Utility Theory (MAUT) provides for multi-dimensional attribute variables. However, the basic postulate remains the same, that the objective function is expected utility.

Consider a space of variables (states of the world), which may include physical states, utilities and decision (information/choice) states, within the model, and some other variables, e.g., cost, schedule, alternatives and utilities outside the model. Some subset of these are called attributes for the particular decision to be made, inside or outside the model. The combination of attribute dimensions represents a joint outcome domain. The utility function over these attributes is a joint function, and there is also a joint probability function over these outcomes. A MAUT with full generality assigns worth to each joint outcome, conditioned on the chosen alternative, multiplies these by the joint conditional probability distribution and sums over the joint outcomes to obtain expected utility conditioned on the choice. If there are n attributes of interest, the attribute space can be represented by $X = \{X_1 \times X_2 \times ... \times X_n\}$ and an outcome in that space would be $(x_1, x_2, ..., x_n)$. Then equation (1) becomes

$$EU(A) = \int u(x_1, x_2, ..., x_n) \bullet p(x_1, x_2, ..., x_n \mid A)\, dx_1\, dx_2 .... dx_n . \tag{2}$$

or, $$EU(A) = \int u(\mathbf{x}) \bullet p(\mathbf{x} \mid A)\, d\mathbf{x},$$

for short, where $\mathbf{x}$ is the vector, $(x_1, x_2, ..., x_n)$.

But this form of the expected utility implies needing to know not only the joint utility function over all the attributes, but also the joint probability distribution for all joint outcomes for each alternative. The feasibility of applying Decision Analysis based on MAUT depends on the nature of the problem. If the domain does not have many outcomes, the problem is simplified. If the outcome is determined exactly (with probability one) by the alternative chosen, then only utilities for those few outcomes need be known. But most problems are more complicated. There are classes of decision problems for which the structure of the utility and probability functions can be exploited to reduce the difficulty. One class is the additive independence case and the other is the utility independence case. Knowing which class applies requires careful assessment of the utility function.

2.2.1 Additive Independence

The most effective case is when the utilities over each attribute, determined separately, are additive to form the utility function over the joint domain. In this case,

$$u(\mathbf{x}) = u(x_1, x_2, ..., x_n) = \sum_i k_i\, u_i(x_i), \tag{3}$$

and, by putting equation (3) in equation (2),

$$EU(A) = \sum_i \int k_i\, u_i(x_i) \bullet p(x_i \mid A)\, dx_i = \sum_i k_i\, EU_i(A), \tag{4}$$

where the k_i are scaling coefficients which bring the joint utility function into the range of zero to one. Note that these are <u>not</u> importance weights on the attributes.

The effect is to cause the calculation to be based only on the marginal distribution, $p(x_i \mid A)$, of each attribute. This is an artifact of the form of the utility function, not the probabilities. It does not matter whether the joint probability distribution is (conditionally) independent with respect to the marginals; it is the utility function that is called "additive independent" in this case. This approach weights the expected utility for each attribute to obtain an overall utility. (The weights are often interpreted to mean relative importance of each attribute to the person making the choice. But this is not necessarily the case; the weights are scaling coefficients to calibrate the utility function.)

2.2.2 Utility Independence (Multiplicative Form)

A weaker effect is that of utility independence. This only results in a form which is multiplicative in the utility functions of the separate attributes, i.e.,

$$k\, u(x_1, x_2, \ldots, x_n) + 1 = \prod_i [\, k\, k_i\, u_i(x_i) + 1], \tag{5}$$

where the k_i are different from those in the additive case, and k is a scaling factor to balance the equation. This form does not make the calculation of the Expected Utility any simpler than the joint utility, unless the distributions of the attributes are conditionally independent of each other, i.e., the joint probability distribution is the product of the marginal distributions; i.e.,

$$p(\mathbf{x} \mid A) = \prod_i p(x_i \mid A). \tag{6}$$

Then,

$$k\, EU(A) + 1 = \prod_i (k\, k_i\, EU_i(A) + 1). \tag{7}$$

Neither of these cases is likely to be satisfied in the types of analyses involving highly complex systems.

2.3 PROXY ATTRIBUTES

When attributes and their utilities are assessed, it may be the case that the attributes are actually proxies for higher level attributes. For example, the attributes of income and net worth may be proxies for the comfort or peace of mind that accrue from that condition. The relationship between the utility of the higher attributes and that of the proxies should depend on the relationship of the attributes themselves.

When a relationship (model) of the conditional effect of the proxy attribute, y, on the higher level attribute, x, is known, a relationship between the utility functions can be found. The expected utility is

$$\begin{aligned} EU_X(A) &= \int u_X(x) \cdot p(x \mid A)\, dx \\ &= \int u_X(x) \cdot \left[\int p(x \mid y, A) \cdot p(y \mid A)\, dy \right] dx \\ &= \int \left[\int u_X(x) \cdot p(x \mid y, A)\, dx \right] \cdot p(y \mid A)\, dy \end{aligned} \tag{8}$$

where $u_X(x)$ is the utility on x, and $p(x \mid y, A)$ is the model of the relationship of x conditioned on y. But for utility on y,

$$EU_Y(A) = \int u_Y(y) \cdot p(y \mid A)\, dy \tag{9}$$

In order for these two expected utilities to be equal, it must be true that

$$u_Y(y) = \int u_X(x) \cdot p(x \mid y, A)\ dx \ \equiv\ u_{X|Y}(y \mid A). \tag{10}$$

The integral will be referred to as the conditional utility function of x given y. This relationship is critical to understanding the relationship between Mission Success Criteria (MSCs) and Required Capabilities (RCs) and whether assessing utility on capabilities rather than MSCs is valid.

Notice that the conditional utility function is, in general, conditioned on the alternatives, unless the model of the attribute relationship is conditionally independent of the alternatives, i.e.,

$$p(x \mid y, A) \equiv p(x \mid y), \tag{11}$$

for all alternatives. Then,

$$u_{X|Y}(y \mid A) \equiv u_{X|Y}(y) \tag{12}$$

is independent of the alternatives.

But utility functions should not depend on the alternatives. This can be a problem in complex models such as those needed to assess military systems. For example, if the alternatives consist of choices among different force levels, but the low level attributes consist of single system capabilities, the model will depend on the alternatives. This can be corrected by treating the force level as an attribute, but does the force level have any intrinsic utility to the decision maker? That is, can a non-constant value or utility function be found for that attribute that is not, in fact, a projection of what the decision maker believes will be the outcome due to the force level?

The proxy attributes, y, above, could be the capability attributes of a system, while the high level attributes of the TLWR could be the x attributes. This is the manner in which the relationship between capabilities and TLWR outcomes can be incorporated into an MAUT analysis.

2.4 DECISION TREES AND INFLUENCE DIAGRAMS

The decision "tree" used in decision analysis refers to a branching of the individual outcomes of a cascade of variables. Each branch of the tree, at each stage of branching, is labeled with an outcome and a probability of that branch occurring, given that the branch point was reached by the series of branchings from the starting point (the decision node being considered). In other words, the probabilities are conditional probabilities of the branch occurring, given the antecedent events between the starting point and the branch point. The order of the stages of the branches is not required to be in the causal order, but that is usually the result when people think about the problem. At the leaf nodes at the end of the tree, utilities are listed for the joint outcome of all the branch occurrences between the starting point and the leaf node. The "roll-up" of the decision tree consists of multiplying the utility at the leaf node by the probabilities on the leaf branches and adding the products for all branches leaving one point for each stage of branching. This results in utility for the branch point. The process is repeated for all the stages of the tree. The effect of this is to multiply the conditional probability distributions together for all stages and then multiplying by the joint utility function. The initial stage of the tree displays the alternatives of the choice and does not have probabilities assigned. Rather, the first branch point, after the roll-up is performed, contains the expected utility for the alternative on that decision branch.

The decision tree is useful for small problems with discrete variables or outcomes. It lets the decision maker see all the outcomes and branches. But it cannot handle continuous variables and their distributions, and for complex problems, the tree can become intractably large. Note that the branching is a string of events, with all possible combinations being accouinted for. This is exactly a chain of conditional outcomes, just as in the causality net (Girard, 1989A). But there is no way, in the decision tree, to account for conditional independence and reduce the complexity of the representation. The influence diagram approach to viewing decision analysis (Howard and Matheson, 1984, and Schachter, 1986) is more like the causality net approach. The influence diagram is a graph of the conditioning relationship between event nodes, including outcomes, values and choices. It also has the basic decision node. It is a more compact view of the decision tree. Just as in the decision tree, the order of the nodes is not required to be in the causal direction, but when they are and all conditional independencies are recognized by removing arcs between conditionally independent variables, the result is a causality net. In the influence diagram, there is a value node, which represents the utility function. The value node is dependent on (has arcs from) those variables which are considered as decision attributes. (Any node in the net that does not contribute to the value node is a variable to which the decision maker is indifferent.) The relationship between the resultant value outcome and the decision alternatives is represented by the influence diagram and the process of solving it is similar to, though more complicated than, that of the causality net, unless they are recognized as the same.

2.5 MEASURES ON VALUE FUNCTIONS

The probability distribution of the value node of the influence diagram or causality net can be calculated, conditioned on the alternatives of the decision node, by

$$p(u \mid A) = \int p(u \mid \mathbf{x}) \cdot p(\mathbf{x} \mid A)\, d\mathbf{x}, \tag{13}$$

where $p(u \mid \mathbf{x})$ is the deterministic probability of the value $u = u(x)$ (where $u(x)$ represents either a utility function or a value function), that is

$$p(u \mid \mathbf{x}) = p(u = u(\mathbf{x})) = 1, \text{ when } u = u(\mathbf{x}), \text{ and } 0, \text{ otherwise.} \tag{14}$$

The value or utility probability distribution, $p(u \mid A)$, contains more information than just the expected utility; it shows the variability of the result. The expected utility can be found by taking the expected value of the distribution of the value node, i.e,

$$EU(A) = \int u \cdot p(u \mid A)\, du. \tag{15}$$

If x_i is a single ordered variable with a measure, we could also calculate the expected value of x_i for each alternative, A, by

$$EX_i(A) = \int x_i \cdot p(x_i \mid A)\, dx_i. \tag{16}$$

But many outcome variables, such as win/lose or red/green/blue or discrete variables, do not have a measure, so their expected value is meaningless.

The equations above show that there are many ways to view the measures of the problem: $EX_i(A)$, $p(x_i \mid A)$, $p(\mathbf{x} \mid A)$, $EU(A)$ or $p(u \mid A)$. Each of these carries progressively more information about the result. MAUT stops at EU(A), whereas $p(u \mid A)$ is also available with the same model and <u>no additional information is required</u>. The utility distribution function is not a part of MAUT as presently constituted.

3.0 APPLYING MAUT TO DECISION MAKING

There are different kinds of decisions to make. For example, there are operational decisions and acquisition decisions. The nature of the decision will affect the worth of the outcomes or the scaling weights of attributes. The operational decision is one of choice of action and choice among resources available. The acquisition decision is a choice of which resources to develop. The costs to the operational decision maker are lives and production costs of losses. The acquisition decision maker may consider those costs, too, but the development cost also is weighed. These costs are additional variables in the analysis. The utility function can be used to put all costs and worth in balance.

A Hierarchy of Objectives (NOSC TD 1938, Vol.1) is a set of relationships among objectives and subobjectives that relates the set of outcomes that result from processes at a lower level to preferred outcomes at a higher level. The whole structure is identical with the cause and effect relationship of the dependence of certain values of attributes on the values of other attributes (or the same attribute at an earlier time). As a method of system description, this structure is called a causality net . A deterministic or stochastic view of this relationship depends only on whether there is uncertainty in the cause and effect relationship. The measure of this uncertainty is described by conditional probability distributions. Aggregation of the likelihood of top level outcomes can be obtained by suitable multiplication and summation of these measures.

A decision maker at a certain level in the Hierarchy of Objectives will have certain attributes that are important at that level; other attributes that are part of the analysis will be a matter of indifference. In particular, the decision maker should be indifferent to the lower level outcomes that are aggregated to realize the outcomes at the higher level, i.e., those of interest to the decision maker. Someone charged with choosing among similar systems at the lower level has the capabilities of the system as their attributes of interest, but for a decision involving many system types and configurations, only high level criteria can be used to discriminate among alternatives. But, to assess a decision at the higher level, it is necessary to have a means to infer the top level outcomes from the capabilities at the lower level. Aggregating probability over a causality net would provide a distribution of the top level outcomes, to which the worth of those outcomes could be applied. But the capabilities of the alternatives being considered must also be stated in terms that can be incorporated into the process.

3.1 EFFECTIVENESS FUNCTIONS

The way to fully represent capabilities is with effectiveness functions. Effectiveness functions are conditional probability functions that represent a model of a process. They are conditioned on the situation and the effects that determine results. For example, Detection (vs. non-detection) is conditioned on a target type at a particular range under particular environmental conditions. The "range" capability is a particular point in the detection effectiveness (probability) function at which the probability (of detection, given target type and noise and range) is 0.5. This point will vary depending on the other conditioning variables, e.g., target type and noise. By this definition of effectiveness function, measures of effectiveness and measures of performance are synonymous (Girard, 1989B), since low level processes and high level processes are both modeled as conditional probabilities, the latter having the low level variables integrated out, after combining the conditional models of performance. (The distinction is no greater than the distinction between Mission Success Criteria and Required Capabilities. The former are the goal objectives (results, dependent variables/measures) while the latter are a possible set of the design parameters (inputs, independent variables/measures).) However, sometimes "performance" is expressed in terms of the conditioning variable, such as range. These need to be converted into causal conditional forms, such as $p(y \mid z, A)$, where the variables in z represent situational conditions, such as range, season or scenario variables. Then equation (8) becomes

$$EU_X(z, A) = \int \left[\int u_X(x) \cdot p(x \mid y, z, A) \; dx \right] \cdot p(y \mid z, A) \, dy, \tag{17}$$

and equation (9) becomes

$$EU_Y(z, A) = \int u_Y(y) \cdot p(y \mid z, A) \, dy. \tag{18}$$

The distribution of utility, equation (13), in this case, becomes

$$p(u \mid z, A) = \int \left[\int p(u \mid x) \cdot p(x \mid y, z, A) \; dx \right] \cdot p(y \mid z, A) \, dy. \tag{19}$$

3.2 CONTEXT CONDITIONS

The decision also depends on the context of the question being asked. To provide for this, the situation, environment and the threat must also be described, as well as the procedures, tactics and organizations that will be assumed and modelled. Since these become antecedents of the conditional model, as noted in equations (17), (18) and (19), there needs to be a weighting, in terms of the likelihood of the various conditions occurring, incorporated in the statement of requirements and in the model calculation. Or the results can be presented as conditional on each class of situation, as in equations (17), (18) and (19). In the case of weighting, equation (17) is converted to

$$\begin{aligned} EU_X(A) &= \int EU_X(z, A) \, p(z) \, dz, \\ &= \int\int \left[\int u_X(x) \cdot p(x \mid y, z, A) \; dx \right] \cdot p(y \mid z, A) \, p(z) \, dy \, dz, \end{aligned} \tag{20}$$

where the dependence on the situation variables, z, has been integrated out, and the utility distribution, equation (19), becomes,

$$\begin{aligned} p(u \mid A) &= \int p(u \mid z, A) \, p(z) \, dz. \\ &= \int\int \left[\int p(u \mid x) \cdot p(x \mid y, z, A) \; dx \right] \cdot p(y \mid z, A) \, p(z) \, dy \, dz. \end{aligned} \tag{21}$$

In fact, $u_X(x)$ and, therefore, $p(u \mid x)$, could also depend on the context variables, z, so that

$$\begin{aligned} EU_X(A) &= \int EU_X(z, A) \, p(z) \, dz, \\ &= \int\int \left[\int u_X(x, z) \cdot p(x \mid y, z, A) \; dx \right] \cdot p(y \mid z, A) \, p(z) \, dy \, dz, \end{aligned} \tag{20'}$$

and,

$$\begin{aligned} p(u \mid A) &= \int p(u \mid z, A) \, p(z) \, dz. \\ &= \int\int \left[\int p(u \mid x, z) \cdot p(x \mid y, z, A) \; dx \right] \cdot p(y \mid z, A) \, p(z) \, dy \, dz. \end{aligned} \tag{21'}$$

In this way, one may recognize that circumstances change the utility of outcomes.

3.3 INTERMEDIATE VARIABLES

In a complex analysis, the relationship between the attributes of worth, x, and the capabilities, y, and situations, z, may also involve some intermediate variables, w. These are state variables which are part of the scenario and the functional evolution of cause and effect events. It is straightforward to expand the utility formulas for incorporating these intermediate variables. Equations (20') and (21') become

$$EU_X(A) = \int\int\int\int u_X(x, z)\, p(x \mid w, y, z, A)\, p(w \mid y, z, A)\, p(y \mid z, A)\, p(z)\, dx\, dw\, dy\, dz, \tag{22}$$

and,

$$p(u \mid A) = \int\int\int\int p(u \mid x, z)\, p(x \mid w, y, z, A)\, p(w \mid y, z, A)\, p(y \mid z, A)\, p(z)\, dx\, dw\, dy\, dz. \tag{23}$$

This is the most general form of the utility functions, which includes all the variables involved in the model of outcomes and capabilities, conditioned on the alternatives.

3.4 ELEMENTS OF ANALYSIS

In summary, an analysis must address the following parts:

(1) Identify the attributes of interest, x, in this case, the Mission Success Criteria.

(2) Identify the situation, environment and threat variables, z, including their assumed capabilities in conditional form. It may also be necessary to determine the weighting of the situations according to their likelihood of occurrence, p(z).

(3) Determine the utility function, $u_X(x, z)$, for those criteria (not just the confidence level). It should also be determined whether the utilities are additive or multiplicative or neither, and whether they depend on the situation.

(4) Identify the system alternatives, A, to be compared and their desired (or expected) capabilities in conditional form, $p(y \mid z, A)$.

(5) Construct a model of the probability of the MSC variables for each alternative, $p(x \mid w, y, z, A)$, and a model of the intermediate variables, $p(w \mid y, z, A)$.

(6) Determine the probability distribution of the utility, $p(u \mid A)$, and the expected utility, $EU_X(A)$, for each alternative.

(7) Choose among the alternatives based on (6).

Items (1), (2), and (3) should be derived as part of the TLWR definition process, with utilities being elicited at that level. The alternatives in (4) are determined by the question being posed, but the capabilities come from system requirements documents or they can be elicited as part of the appraisal. However, for C3 systems, the issue is that the nature of the capabilities that are important to the analysis has not been agreed upon. The Decision Probability (NOSC TD 1938, Vol.2) has been proposed as the equivalent of the Detection Probability and Kill Probability. Item (5) is the most difficult part of the process. The model can be exceedingly complex, but achievable in principle. The most difficult part of the model is incorporating the decision probability into the model. In order to do this, a good descriptive model of decision making is required, which, to date, has not been accomplished. If a good model existed, steps (6) and (7) would be straightforward.

Lacking the good model, the alternative method would be to characterize the system alternatives in terms of their capabilities and determine the utility function on these capabilities. In this case, the capabilities become the "proxy attributes" (see Section 2.3). But it is not likely that the utility functions assessed on these proxies would bear any resemblence to the conditional utility function based on the model and the high level attributes of the Mission Success Criteria. This is the fundamental dilemma of conducting appraisals that attempt to relate capabilities to MSCs. Either the utility function on the MSCs cannot be used because the model is not known or the utility function on proxy attributes can not be validated to represent the conditional utility function, because the model is not known. What is needed is an efficient, reasonable model, which reflects the effects of operational decision making as well as the system capabilities and force structures, in order that the acquisition decision making process can be executed.

4.0 ADDITIONAL METHODS OF DECISION ANALYSIS

4.1 IMPRECISELY SPECIFIED MULTIPLE ATTRIBUTE UTILITY THEORY

Imprecisely Specified MAUT (ISMAUT) is a method of continuing the analysis even when the decision makers are unable to precisely specify their utility and probability statements. The decision maker may say that the weight or probability lies above or below a particular level or between two levels (interval specification) or that one of the weights or probabilities is more or less than another (cardinal or ordinal ranking). These statements, and the requirement that weights and probabilities must add to one, can be used in a simple linear programming approach to find a region of feasible combinations of the weights or probabilities that satisfy these specifications, if they are not inconsistent (having no feasibility region). Since the regions are intervals in weight space, there may not be a unique ordering of the alternatives as a result.

4.2 FUZZY DECISION ANALYSIS

Fuzzy Decision Analysis (FDA) attempts to use fuzzy set logic to characterize the decision problem. Rather than expected utility as a criterion, FDA uses a utility membership function formed by combining membership functions of utility and probability using the fuzzy operators "max" for summation and "min" in place of products. The operators are applied to membership functions representing the degree to which an outcome has utility for various outcomes and the degree to which a particular probability of that outcome is possible. These terms do not provide a clear notion of the meaning of these membership functions and the concepts are difficult to understand. The membership functions can be viewed as fuzzier versions of the intervals used in ISMAUT, where the degree to which a weight or probability belongs to the interval is one in the interval and zero outside of it. The fuzzy interval has membership values between zero and one. There is even less information purported to exist in the fuzzy interval than in the crisp intervals of ISMAUT, so one should expect not to be able to draw any surer conclusions. When there is less information known about the data than required by the MAUT method, some fall-back position must be taken. The ISMAUT approach only relaxes the specification requirement, but follows the expected utility paradigm. The FDA approach further relaxes the specification requirement, but it also changes the method of calculation. Other work (Cheeseman, 1986) shows that fuzzy sets can be viewed as conditional probabilities. The membership functions are, in effect, conditional probabilities that the utilities and probabilities fall into some range of values. This suggests that a probability that the overall expected utility falls into some interval can be calculated using probability operators rather than fuzzy operators. This is suggested for further research.

4.3 ANALYTIC HIERARCHICAL PROCESS (AHP)

In AHP, objectives, functions and capabilities are decomposed in a hierarchical manner. A measure of "influence" of these attributes on one another is obtained from the decision maker. This influence sounds like it could be interpreted as the subjective conditional probability of one attribute given another. For a causality net, the conditional probability must be the influence of the joint state of all antecedent attributes on the outcome of the dependent attribute, so the AHP approach would not reflect the proper aggregation of these influences.

Another aspect of the AHP is the comparison of alternatives, pair-wise, with regard to each "criterion", using a ratio scale. Each pair, when compared in reverse order, should exhibit an inverse ratio to the original comparison, for consistency. The matrix of comparisons for each criterion is analyzed to determine the eigenvector of its largest eigenvalue. This determines a score for each alternative for each criterion. The composite score is determined by the influence hierarchy and the alternatives are compared on the basis of this score.

Significant problems have been noted concerning AHP, the most crucial being that adding an alternative with capabilities near one of the other alternatives can change the order of preference of the other alternatives. A solution to this for special cases is recommended by Dyer (in Winkler, 1990). The adjustment is to scale the eigenvectors' values to make them look like utilities by a linear transformation (subtracting the smallest and dividing by the largest (after subtracting)). When this is done the result is compatible with MAUT for additive utilities and direct assessment of attributes (without aggregating up a hierarchy), but AHP does not correspond to the general MAUT solution for non-additive utilities and proxy attributes.

4.4 SUBJECTIVE TRANSFER FUNCTION APPROACH

The Subjective Transfer Function (STF) method also generates a hierarchy, generally a tree, of objectives and outcomes and their influences. The STF approach supposes that people can judge the outcome level of an objective, given some level of capability of its antecedents in various areas. The relationship between the outcome and the antecedents is hypothesized to be from a set of algebraic functions. Questions are used to elicit responses for various joint combinations of antecedent levels and a fit to the set of algebraic functions is attempted, including fitting weights. The"best" fit is used as a subjective transfer function for that branching point of the tree.

While the STF method recognizes that there are multiple influences for outcomes, there is no basis for not allowing a low tier attribute from influencing a higher level one, which forces a tree-like structure. The elicited level of the outcome could be related to a subjective conditional probability, but there is no evidence for this. The STF method has no objective function except for maximizing the top-most attribute, which is not stated to be any form of utility.

4.5 HIERARCHICAL VALUATED STATE SPACE

The method of Hierarchical Valuated State Space (HVSS) decomposes a set of objectives and characteristics of a problem in a tree-like fashion, until, at some level, "goodness" values, depending on the alternative under consideration, are assigned to the leaf nodes. At each branch point, the branches are given relative importance weights that add up to 1. The "roll-up" consists of multiplying the values by the weights to get a value for the branch points, until the root node of the tree is reached. The results are compared for the various alternatives. If the leaf nodes were, in fact, separate attributes, the tree would look like a tree of importance weights in MAUT, which is a special view of the attributes, not a decision tree. In MAUT, the importance tree is only a construct formed by combining attributes into groups, and the weights are the importance within the group. The groups have no relation to objectives. It only applied when the attributes were additive independent. If it was known that additive independence held, then the weights in the additve utility function were known. The importance tree was only proposed as a way to find out the relative weights of the attributes when there were a large number of them. Other methods for determining the weights were used. Furthermore, if there are too many attributes to assess, then, possibly, most of them should be lower level attributes to which the decision maker should be indifferent. The decision analysis will fail unless the truly relevant few attributes of importance are identified. The lower level variables may still be important to the analysis, but only as intermediate variables, such as capability dimensions or situations.

The HVSS approach is a misinterpretation of the importance tree noted by Keeney and Raiffa (1976). Keeney and Raiffa also note that the scaling factors in the additive independence case cannot be interpreted as importance weights. The HVSS does not have an objective function, other than the hierarchical sum of the weighted goodness values. This sum has no interpretation in terms of the outcome likelihood or utility of the attributes.

4.6 JOINT DECISION SUPPORT SYSTEM

The Joint Decision Support System (JDSS) approach provides a way to construct alternative C3 system configurations and to assess them. The assessment method decomposes the problem, for each Unified Commander-in-Chief (CINC), into Warfighting Environment (situation), Major Mission Area (e.g., Maritime), Mission Element (e.g., ASW), Level of Capability, Functional Task (e.g., Monitor), and Capability Objectives. Importance of each of the elements at these levels with respect to the part of the next higher level are entered in verbal (categorical) terms. An assessment of capabilities is made at the lowest level and rolled up to any level of the hierarchy using ad hoc combination rules.

The JDSS was supposed to apply fuzzy logic for the aggregation of the assessment, but the technique is not implemented in that way. The hierarchy used is a way of organizing data, not a cause and effect relationship. There are no high level objectives or attributes. "Capability Objectives" are low-level attribute descriptions with parameter values as goals sometimes stated. An assessment of an alternative C3 system configuration bears no relation to warfighting objectives, because they are not included in the assessment. Only C3 functions are represented.

The JDSS decomposition is similar to that of a HVSS approach except that JDSS uses words and HVSS uses numbers for the valuation, importance, and roll-up process.

5.0 STRUCTURING AN MAUT PROBLEM

5.1 GENERAL APPROACH

As discussed in Section 3.4, an MAUT analysis consists of identifying the attributes of interest, the utility of the attributes, the situation variables, the alternative system configurations, their capabilities and the intermediate variables involved in modeling the attribute outcomes. These six elements of the method are depicted in Figure 5-1.

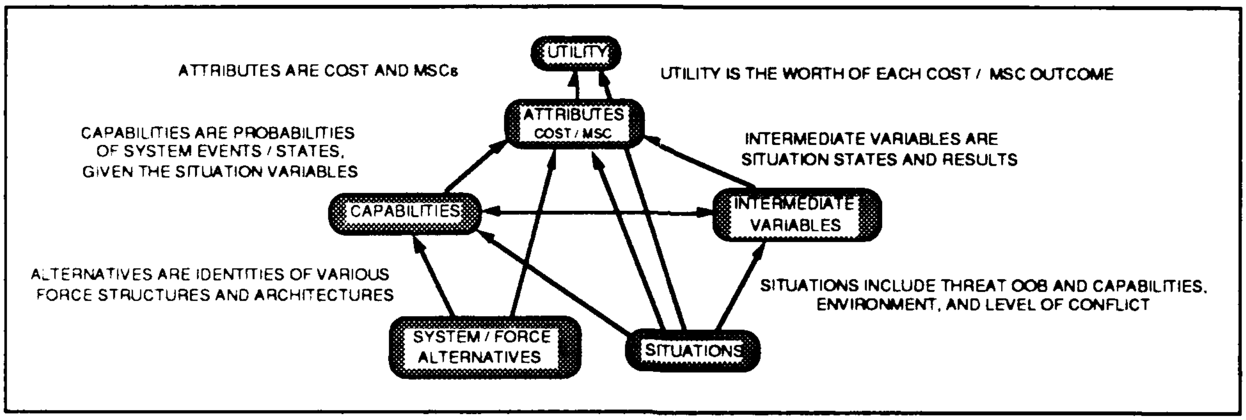


Figure 5-1. Elements of Extended MAUT Approach

This approach is an extension of MAUT because it inserts the modeling of outcomes into the process, so that utilities are not applied to the characteristics of the alternatives, but to the attributes of interest. The model bridges the gap between the alternatives (and their capabilities) and the outcomes of the attributes of interest, providing the relationship between them. It also makes explicit the dependence on the situation and requires intermediate variables for completeness.

The utility of each alternative is a combination of the probabilities of the attributes, given the alternative is implemented, and the worth function on the attributes. The worth of the attributes is a TLWR issue, which may also depend on the situation. The probabilities of the attributes are produced by the model, with dependence on the situation, alternatives, capabilities and intermediate variables, including the enemy's capabilities.

Some examples of these elements are:

(1) Attribute of importance (Mission Success Criteria) = Number of units remaining. This integer variable runs from zero to the number of units, N, available in the force considered, so it depends on the alternative (or the situation variable of enemy units available, if the MSC is the number of enemy units remaining). If the number of units available is not different among the alternatives, then it is a situation variable.

(2) The utility of the attribute might be u(0) = 0, u(1) = 0, u(2) = 0, u(3) = 0, u(4) = 1, u(5) = 1, meaning only the last two outcomes are good. Another utility function might be u(0) = 0, u(1) = 0, u(2) = 0.2, u(3) = 0.6, u(4) = 0.9, u(5) = 1. Note that this utility function looks like a fuzzy set of the concept of a "good" outcome. These utility functions might represent "good" in different situations. This dependence is included in the form of the evaluation (equations (22) and (23)), where u(x, z) depends on the situation, z.

(3) The probability of the attribute outcome for a particular alternative might be $p(0 \mid A1) = 0.05$, $p(1 \mid A1) = 0.08$, $p(2 \mid A1) = 0.1$, $p(3 \mid A1) = 0.17$, $p(4 \mid A1) = 0.4$, $p(5 \mid A1) = 0.2$. The expected utility of this outcome would be 0.6 for the first utility function in (2), and 0.684 for the second one. The point is that without a change in the probabilities, the expected utility is different because the worth of the low number outcomes was higher in the latter case of the utility function, so the worth function is an important part of the analysis. Note that the confidence that $n \geq 4$ is 0.6, also. The first utility function in (2) is also a mechanism for computing this threshold confidence factor for $n \geq 4$. The general form of a threshold confidence measure has the utility function equal to one for outcomes considered "good" and zero for "bad", in which case, the confidence is the sum of probabilities for the good cases. The fuzzy form is just as useful a form of the measure. These are identical ideas, but in the latter case, the good/bad concept is not as black and white. It provides a means to incorporate less valued outcomes into the analysis, without giving them full weight or zero.

The distribution of utility can also be calculated for these two utility functions. In the first case, $p(u = 0) = 0.4$ and $p(u = 1) = 0.6$. In the second case, $p(u = 0) = 0.13$, the sum of the two probabilities for which $u = 0$, i.e., $p(0 \mid A1) + p(1 \mid A1) = 0.05 + 0.08$. Since each of the other values of utility, in the second case, are distinct, their probabilities are equal to the probabilities of the corresponding attribute outcomes: $p(u = 0.2) = p(2 \mid A1) = 0.1$, $p(u = 0.6) = p(3 \mid A1) = 0.17$, etc.

(4) A capability statement is, in effect, an effectiveness function as described in Section 3.1. The example given was that of the detection effectiveness (probability) function, which is the probability of detection, given target type and noise and range, for the given system alternative. Here "target type" is a situation variable; "noise" is an intermediate variable, which depends on the situation variable, environment; "range" is an intermediate variable, which depends on the dynamics of the scenario; and the "system alternative" is a member of the alternative set.

These elements are further exemplified by the following EW problem.

5.2 AN EXAMPLE OF MAUT APPLIED TO EW

Consider an Electronic Warfare (EW) system problem, consisting of sensors (electronic support measures (ESM)) and electronic countermeasures (ECM). The <u>mission</u> (purpose) of the system is to prevent damage. The <u>functions</u> of the system are to detect, characterize (identify platform and emitter types), and localize sources of electronic emissions, to cue countermeasures when a threatening situation occurs, and to decide to deceive, decoy or jam the threat and carry out the decision. Outside this system, there may also be hard-kill weapons to be cued and used to destroy the threat. The <u>tasks</u> of the sensors include determining direction, elevation, frequency/band, PRI, pulse width, scan rate, modulation type, or other unique characteristics. The sensor systems must also determine the likelihood of the emission being from a particular platform or weapon type. Cueing the countermeasures or weapons is a task of the decision function, which involves notifying them of the known characteristics and parameters. The decision function must also decide to cue and what action to take, based on the nature of the situation, operational procedures and rules of engagement in force. The task of the soft- and hard-kill weapons is to prevent being hit. Finally, the enemy weapon causes any damage that occurs, but only to the degree that damage control can prevent it, i.e., damage control is a causal limitation on damage.

The attributes of this problem are the amount of damage incurred (Mission Success Criterion) and the cost of the EW system. An operational model is needed to determine the probability of damage and a cost model is needed to determine the probability distribution of the cost. The cost model may include development cost, production cost and logistics cost.

Figure 5-2 presents a causality net of the dimensions of this problem. The cost model is on the left and the operational model constitutes the majority of the figure. Arrows arriving at a dimension indicate which dimensions effect it. This is the causal conditioning relationship, "state X (at the head of the arrow) is conditioned on (or depends on) the states, Y (at the tails of the arrows leading to it)." When there is a probability associated with this conditioning, it is interpreted as the probability of X, given the joint state of the variables, Y, leading to it. It is useful to consider all conditional relations as being "probabilities" for the purposes of being able to manipulate equations of the relationships. When the relationship is deterministic, the probability used is 1.0, for the case in point, and 0.0 for all other outcomes, given the joint state of the variables leading to X. This is why, for example, in equations (1), (2), (8), (9) and (17), the probabilities of the outcomes or the capabilities were conditioned on the alternatives. This made explicit the dependence on the alternative in question. The capabilities attributed to a particular alternative can and should be stated in terms of the conditional probability of the capability dimension, such as "parameter values perceived", given parameter values emitted, range, bearing, detection declared, which are scenario variables, as well as the alternative of interest.

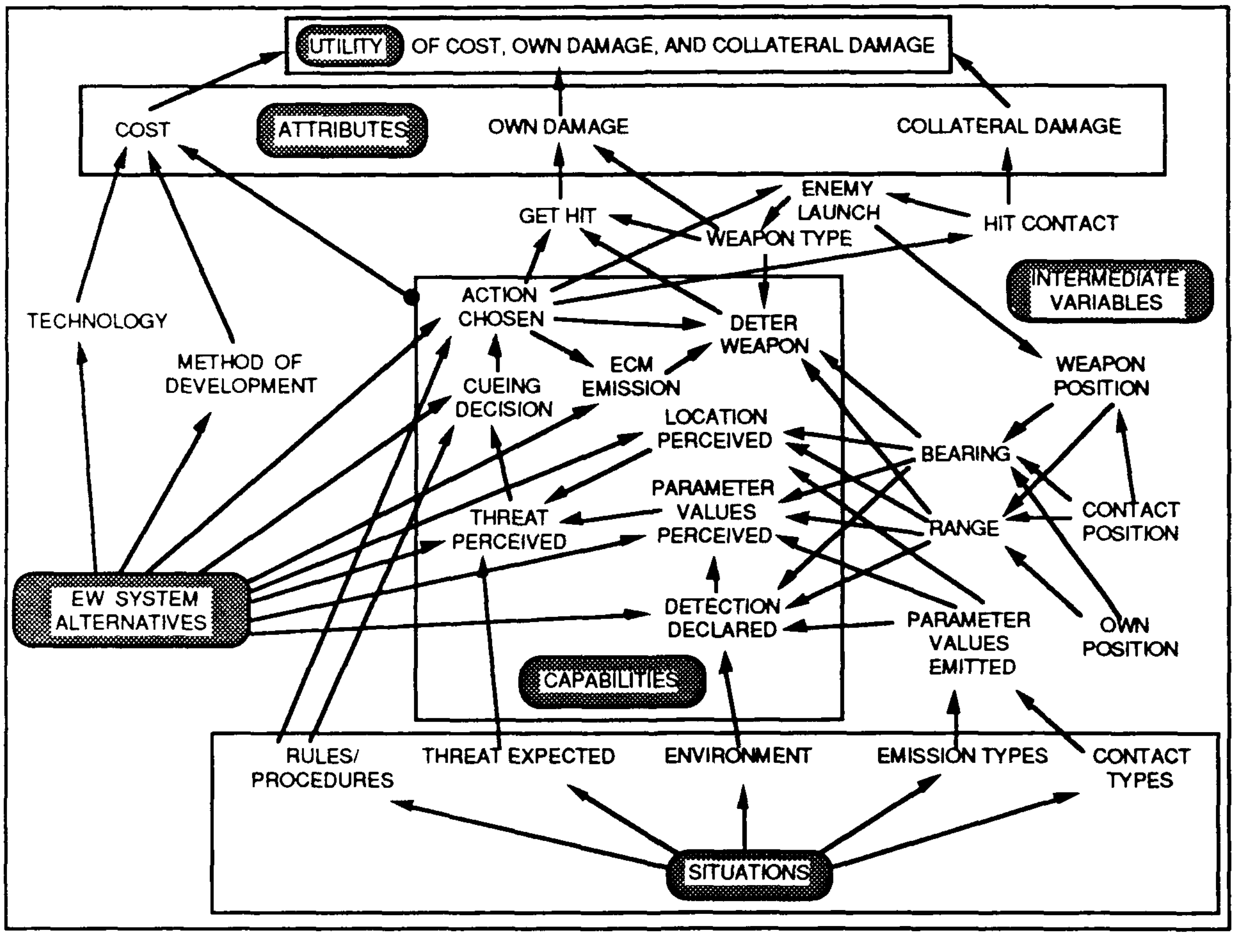


Figure 5-2. Causality Net for EW Example

The operational model involves the probability of being damaged, given a hit by a particular weapon type. The probability of being hit depends on (is conditioned by) the probability of the type of enemy weapon, its being deterred (destroyed or diverted) or other evasive action taken. The deterrence depends not only on the EW system, and the range and bearing of the encounter, but also on what other systems can be brought to bear during the engagement. These other systems may cause the launch platform to be destroyed or prevented from launching. The other systems may be presumed to be the same for all choices of EW system under consideration, but the better they perform, the less valuable any EW system will be judged in the analysis. This is important because the alternative of no EW system should be considered, not just for the purpose of an honest assessment, but also to determine the baseline for the utility on cost. The contribution of the EW system is the change in probability of damage when the EW system is included in the analysis. This includes the possibly increased effectiveness of other weapons due to better cueing, as well as the direct effects of the ECM in the EW system.

The capabilities of the EW system include a large set of conditional probabilities. First, there is the probability of detection, given each platform and emitter type, conditioned also on whether there is an emission, as well as on the range and bearing, and atmospheric conditions. In addition, there is a probability of false alarm for the case when there is no emission. Next, there are the probabilities of identifying each parameter with some value, given there is a detection and the value of the actual parameter. For example, the probability of believing pulse repetition interval (PRI) to be x, given that it is y, depends on the emitter type, and, in particular, pulse shape. Then there is the probability of identifying the platform or weapon type as one kind, given that the system believes the values of the parameters to be some combination. This does not depend on the actual emitter type directly, only on what the system or operator believes about the parameters. But it may also depend on what the operator (or equipment setting) expects the threat to be. An inference mechanism may be biasing the perceived threat due to expectations.

The next aspect of the chain of events involves what to do with the belief that there is a particular type of emission. First, the process may cue other decision processes, which constitutes a dimension of courses of action. These lead to consequences of the courses of action, such as destroying the contact, whether friendly, neutral or enemy, or preventing its weapon from achieving its objective of hitting the protected asset. (Note that the possibility of reacting to a friendly or neutral by mistake suggests an additional mission objective of avoiding collateral damage to friendly or neutral units. This is an additional attribute, not previously considered. The additional attribute of damage to the enemy may also have value, to which the EW system may contribute through its cues.) If the enemy weapon achieves its objective, the probability of damage will depend on the weapon type and the vulnerability of the platform. Actually, the vulnerability of the platform is a conditional probability of degree of damage, given a hit by a particular type of weapon.

The whole operational model is conditioned on the set of situations of concern to the decision makers. The situations are characterized by the environment, the enemy and friendly order of battle which may become contacts, their possible emissions, the threat expected in the operation plan, and the rules of engagement and standard operating procedures asserted to apply to that situation.

The set of alternatives (sometimes called candidate architectures) points to the sets of capabilities that characterize them. The cost model will need to consider all the capabilities as contributors to the differences in the cost comparison, just as the operational model does. But the method of development, the technology of choice in the development and other factors may also come into play in determining cost of each alternative.

The assessment proceeds by calculating the probability of the outcomes of the cost and damage attributes and then applying the utility function to either calculate the expected utility, EU(A), or the probability distribution of the utility, $p(u \mid A)$.

This example demonstrates how difficult it is to state the conditions under which the assessment is to be made and the attributes that are important. The probability of damaging a friendly or neutral depends on how many there are and of which type and how closely they resemble the enemy. The resemblance may depend on different parameters for each confusable type. Therefore, it may be that a discrimination capability for one parameter is important for one combination of enemy vs. friendly, while another parameter may be important for another combination. This illustrates that there is no intrinsic goodness of the discrimination capability for a given parameter, in terms of utility. It depends on the situation. However, there is an intrinsic effectiveness function, which is the conditional probability of discriminating among values of a parameter.

The question is whether we can and should model these effects for the decision making team in order to ensure everyone has the same information about the situations and their outcomes. Otherwise, how can we expect that eliciting some vague numbers about how "well" a system discriminates and how valuable that is in terms of mission success can be useful in comparing alternatives? Furthermore, how can one believe that the weighted hierarchical sum of these "goodness" numbers, as proposed by several alternative methods, bears any relationship to the aggregate probability of attribute results (mission outcome) and the expected utility of those results? There is no justification for these other approaches to decision making under uncertainty. The key word is "uncertainty". Probability is the way to express uncertainty. MAUT is the only method that incorporates probability as a basic constituent.

6.0 CONCLUSIONS

The principal conclusion is that MAUT is the only well-founded method of aiding decision making on a mathematical basis. However, in practice, it may not be tractable, using existing tools, to apply it to the highly complex problem posed by the Director of Naval Warfare, which is how to relate system capabilities to TLWRs. Many other mathematical methods exist, but they only give the illusion of logical rationale. Others are attempts to accommodate the fact that exact probabilities or utilities may not be available either objectively (empirically) or subjectively (mental impressions). ISMAUT is the only method that preserves the axioms of MAUT well, while it should provide a result which is not necessarily ordered, i.e., alternatives may have overlapping ranges of expected utility. This is a result of the uncertainty of the information, not a flaw in the method. Attempts to use fuzzy sets and logic also address the issue of uncertain information, but there is not a good base of research on the justification for techniques of applying it. MAUT, when applied to a manageable problem, such as a single system type with well defined attributes that are truly important to the purpose of the system, may be tractable, if the values of probability and utility are determinable.

When MAUT or ISMAUT cannot be employed, a non-mathematical approach would be better than pursuing a complicated procedure that is all form and has no logical consistency for its conclusions. A non-mathematical approach would entail describing the attributes of the system to true experts in their use, with a facilitator probing to find why various attributes are important. The discussion and the conscious consideration of the nature of the problem is often the most important part of the appraisal process.

Tree hierarchies of importance weights should be avoided. There is a difference between decomposing something into its constituent parts and structuring the functional relationship among those parts. The first results in a tree-like structure; the latter forms a lattice of interdependence. Approaches which use tree hierarchies of importance weights cannot compute expected utility of high level attributes (Mission Success Criteria) from low level capability statements, while a model of the causal conditioning would produce those results.

Approaches which apply utility to the bottom of a decomposition of attributes, should test to see that those utilities are not dependent on the alternatives and that the dependence of high level attributes on the low-level attributes is through the conditional probability model. The dependence on the alternatives may come about when trading off disparate system types. The utility of low-level capabilities of one system may be different, depending on the capabilities of other systems in the various alternative force combinations. The latter dependence test is equivalent to applying utility at the top-level, anyway, and should be the method of performing the analysis, since it is both alternative independent and top-level oriented with respect to utility.

Finally, the combat modeling problem needs to be solved in a way that incorporates the effects of decision support systems. C3 systems have no intrinsic value outside their contribution to the Mission Success Criteria. These effects can only be represented by the basic result of the decision making process, which is the enabling of all other warfare functions.

7.0 RECOMMENDATIONS

The principal recommendation is to develop the appraisal process as a complete MAUT analysis, as described in Section 3.4; but that is probably not affordable. Therefore, this recommendation will be weakened after a discussion of what is entailed by a complete analysis.

The complete MAUT approach, needed to address the Naval Warfare appraisal process, requires a set of attributes (TLWR mission states) and their worth to the decision makers. These should be provided by the mission experts who specify the TLWRs. Another requirement is a set of capability states of systems, which constitute their nature, as well as the conditional probability functions of those capabilities, given the governing conditions for their relevance. Additionally, a distribution of these governing conditions, which constitutes a scenario, must be provided. Finally, a model of the process which turns capabilities into mission outcomes is needed to make the connection. A cost and acquisition model must also be provided, which accounts for the likelihood of those capabilities being achieved and the cost of achieving them. The utility of the alternatives would be compared on the basis of the top level criteria, including cost.

The analysis should be done at the TLWR level, which means that it supercedes the mission of OP-76. The overall assessment issue is an OP-07 responsibility, while the decision making aspects (the most difficult to analyze) are for OP-76 to resolve. OP-76 would be responsible for identifying the capabilities of proposed C3/EW systems and the effectiveness functions for those capabilities. This is a problem which cannot be accomplished out of context of the Warfare Mission Areas and the scenario situations, since the principal contribution of C3 systems is to support decisions which enable the capabilities of the other systems.

Specific Recommendation: Conduct non-mathematical analysis of C3 alternative configurations, as noted in Section 6.0, when it is not feasible to embed the analysis in a multi-warfare assessment using MAUT. There are no attributes of C3 systems that allow a stand-alone assessment of their utility or probability of occurrence, therefore MAUT cannot be used to determine utility with respect to MSCs, without a sound model of the process which incorporates decision making outcomes. No other mathematical method has a foundation to be used for this purpose.

Specific Recommendation: Solicit support from the Office of Naval Research to address the question of what C3 system characteristics improve the probability of recognizing situations and taking appropriate action.

Specific Recommendation: Translate system characteristics and performance into conditional logic format and build a data base of performance.

Specific Recommendation: Obtain measures of worth on the outcomes in the TLWRs.

Specific Recommendation: Define decision outcomes that effect force performance and the information necessary for these decisions. Conduct wargame experiments to observe situations.

Specific Recommendation: Identify the need to incorporate decision probability in all combat models, and the need for submodels of the decision process necessary to generate those probabilities.

8.0 REFERENCES

Andreadakis, S.K., and Levis, A.H., *Accuracy and Timeliness in Decision-Making Organizations* (LIDS-P-1650) Cambridge, MA: MIT, February, 1987.

Barclay, S., Brown, R.V., Kelly, C.W. III, Peterson, C., Phillip, L.D., and Seliridge, J., *Handbook for Decision Analysis.* Technical Report TR 77-6-30. Decisions and Designs, Inc., McLean, VA: September 1977.

Bell, David, E., Raiffa, H. and Tversky, A., Descriptive, Normative, and Prescriptive Interactions in Decision Making, in *Decision Making: Descriptive, Normative, and Prescriptive Interactions.* Eds., New York, Cambridge University Press, 1988, pp. 9-32.

Bresnick, T.A., Annis, C.P., Buede, D.M., *Threat and Requirements Definition, U.S.M.C. Mobile Protected Weapons System.* Proceedings 48 Military Operations Research Society, December 1981.

Brown, R.V., Kahr, A.S., and Peterson, C.R., *Decision Analysis for the Manager.* NY: Holt, Rinehart, and Winston, 1974.

Call, Hollis. J. and Miller, W. A., A Comparison of Approaches and Implementation for Automating Decision Analysis, *Reliability Engineering and System Safety, 30,* 1990, 115-162.

Cheeseman, P., Probabilistic versus Fuzzy Reasoning, in L. N. Kanal and J. F. Lemmer, Eds., *Uncertainty in Artificial Intelligence,* Elsevier Science Publishers (North-Holland), 1986.

Command and Control System Analysis, Naval Ocean Systems Center Technical Document 1938, September 1990, Vol. 1 The Hierarchy of Objectives Approach.

Command and Control System Analysis, Naval Ocean Systems Center Technical Document 1938, September 1990, Vol. 2 Measuring C2 Effectiveness with Decision Probability.

Command and Control System Analysis, Naval Ocean Systems Center Technical Document 1938, September 1990, Vol. C2 System Functions in the Hierarchy of Objectives.

Command Process Model, Naval Ocean Systems Center Technical Document 1937, September 1990.

DeFinneti, B., Recent Suggestions for the Reconciliation of Theories of Probability, in J. Neyman (Ed.) *Proceedings of the second Berkeley Symposium on mathematical statistics and probability.* Berkeley University of California Press, 1951.

Dockery, J., *Decision Support for Acquisition of C^3 Systems.* Submitted to Aerospace American, March 21, 1990.

Dockery, J., C^3 Managerial Decision Aids for Dealing with Uncertainty. *Journal of Statistical Planning and Inference, 20* (1988), pp. 429-452. North Holland.

Dyer, J.S., Remarks on the Analytic Hierarchy Process, *Management Science,* Vol. *36,* No. 3, March 1990, 249-258.

Dyer, J.S., A Clarification of "Remarks on the Analytic Hierarchy Process", *Management Science*, Vol. *36*, No. 3, March 1990, 274-275.

Edwards, W., Use of Multiattribute Utility Measurement for Social Decision Making. In D.E. Bell, R.L. Keeney, and H. Raiffa. *Conflicting Objectives in Decision.* NY: John Wiley and sons, 1977, 247-276.

Fishburn, P.C., Foundations of Decision Analysis Along the Way. *Management Science*, Vol. *35, #4*, April 1989, pp. 387-405.

Freeling, A.N.S., *Possibilities Versus Fuzzy Probabilities - Two Alternative Decision Aids.* Technical Report, Decision Science Consortium, Inc. Falls Church, VA: November, 1980(A).

Freeling, A.N.S., *Fuzzy Sets and Decision Analysis,* IEEE Transactions on Systems, Man and Cybernetics, vol. SMC-10, No.7, July 1980(B).

Girard, P.E., *A Combat and Decision Model Based on Conditional Probability Logic.* Proceedings of the Joint Directors of Laboratories C2 Research Symposium, 1989(A).

Girard, P.E., A Function-Based Definition of (C^2) Measures of Effectiveness, in *Science of Command and Control: Part II Coping with Complexity.* Stuart Johnson and Alexander M. Levis Eds. AFCEA International Press, 1989(B). Fairfax, VA pp. 112-121.

Girard, P.E., *Stochastic Conditional Object Oriented System Analysis.* Proceedings of the Joint Directors of Laboratories C2 Research Symposium, June 12-14, 1990.

Howard, R.A., and Matheson, J.E., "Influence Diagrams", 1981, in Howard, R.A., and J.E. Matheson (Eds.), *The Principles and Applications of Decision Analysis, Vol II,* Strategic Decisions Group, Menlo Park, CA, 1984.

Howard, R.A., "Knowledge Maps", *Management Science*, 35 (1989), 903-922.

Harker, P.T., and Vargas, L. G., Reply to "Remarks on the Analytic Hierarchy Process" by J.S. Dyer. *Management Science*, Vol. *36*, No. 3, March 1990, 269-273.

Harker, P. T., and Vargas, L. G., The Theory of Ratio Scale Estimation: Saaty's Analytic Hierarchy Process, *Management Science*, Vol. *33*, No. 11, November 1987, 1383-1403.

Hays, M.L., O'Connor, M.F., and Peterson, C.R., *An Application of Multi-Attribute Utility Theory: Design to Cost Evaluation of the U.S. Navy's Electronic Warfare System*, DT/TR-75-3, Decisions and Designs, Inc., 1975, McLean, VA.

HEAT Observer Training Workbook, Volume I, Alphatech, Inc., Burlington, MA, and Electrospace, Inc., Arlington, VA, SP 1027.

Hershey, J.C., Kumreuther, H.C., and Shoemacher, P.J., Sources of Bias in Assessment Proceedings for Utility Functions in *Decision Making: Descriptive, Normative, and Prescriptive Interactions.* Bell, D., Raiffa, H., and Tversky, A., Eds. Cambridge University Press, Cambridge, 1988, pp. 422-441.

JDSS Programmer's Manual (Version 2.2), Washington, DC: Defense Communications Agency, November 1989.

JDSS User's Manual (Version 2.2), Washington, DC: Defense Communications Agency, November 1989.

Jin, V.Y., Levis, A.H., and Remy, P., *Delays in Acyclical Distributed Decision Making Organizations*. Proceedings of Fourth AFAL Symposium on Large Scale Systems: Theory and Applications, Zurich, Switzerland: 1986.

Kahneman, D., Slovic, P., and Tversky, A., (Eds) *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press, Cambridge, 1982.

Keeney, R.L., Decision Analysis: An Overview. *Operations Research*, Vol. *30 #5*, September-October 1982, pp. 803-838.

Keeney, R.L., Measurement Scales for Quantifying Attributes. *Behavioral Science*, Vol. *26*, 1981, pp. 29-36.

Keeney, R.L., and Raiffa, H., *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. NY: John Wiley and Sons, 1976.

Larsen, R.W. and Dillard, R.A., *A Fuzzy Logic Decision Support Tool*. Technical Report 1286. Naval Ocean Systems Center, San Diego, CA: October 1989.

Levis, A.H. and Boetecher, K.L., Modeling and Analysis of Teams of Interacting Decision Makers with Bounded Rationality. *Automatica*, 1983, *9(6)*, pp. 703-709.

Levis, A.H. and Monguillet, J.M., *Modeling and Evaluation of Variable Structure Command and Control Organizations*. (LIDS-P-1795) Cambridge, MA: MIT, July 1988.

Louvet, A.C., Casey, J.T., and Levis, A.H., Experimental Investigation of the Bounded Rationality Constraint. S.E. Johnson and A.H. Levis (Eds). *Science of Command and Control*, Washington, DC: AFCEA International Press, 1988.

Luce, R.D., and Krantz, D.H., Conditional Expected Utility. *Econometrica*, 1971, *39* , pp. 253-271.

March, J.G., Bounded Rationality, Ambiguity, and the Engineering of Choice. *Bell Journal of Economics*, 1978, *9*, pp. 589-608.

Miller, G., The Magical Number Seven, Plus or Minus Two. *Psychological Review*, 1956, *63*, pp. 81-87.

NOSC Technical Document 1938, See *Command and Control System Analysis*

NOSC Technical Document 1937, See *Command Process Model*

O'Connor, M. F., Planning for Integrated System Evaluation: An application to SDI, *Science of Command and Control: Part II: Coping with Complexity*, Stuart Johnson and Alexander Levis Eds., AFCEA International Press, 1989, pp. 101-111.

Raiffa, H., *Preferences for Multi-attributed Alternatives* (RM-5868-DOT/RC). Santa Monica, CA: The Rand Corporation, 1969.

Saaty, T.L., An Exposition on the AHP in Reply to the Paper "Remarks in the Analytic Hierarchy Process", *Management Science*, Vol. *36*, No. 3, March 1990, 259-268.

Saaty, T.L., *The Analytic Hierarchy Process*, McGraw Hill, New York, 1980.

Saaty, T.L., A Scaling Method for Priorities in Hierarchical Structures, *Journal of Mathematical Psychology*, 1977, Vol. *15*, pp. 234-281.

Savage, L.J., *The Foundations of Statistics*, John Wiley, New York, 1980.

Shachter, R., Evaluating Influence Diagrams, *Operations Research, 34*, 1986, 871-882.

Shachter, R.D., and Kenley, C. R., "Gaussian Influence Diagrams", *Management Science*, 35 (1989), 527-550.

Simon, H.A., A Behavioral Model of Rational Choice. *Quarterly Journal of Economics*, 1978, *9*, pp. 589-608.

Small, A.W., and Groveston, P.K. *Strategic Defense Battle Operations Framework*, MTR-85W00114, The MITRE Corporation, July 1985, McLean, VA.

Stewart, B.S., Scherer, W.T., Wykes, E.A., and White, C.C. Defense Communications Decision Support Using ISMAUT and Extensions. Charlottesville, VA: Intelligent Systems Corp., n.d. (circa 1990).

Tversky, A., and Kahneman, D., Judgment Under Uncertainty: Heuristics and biases, *Science*, 1974, *185*, pp. 1124-1131.

Veit, C.T., Callero, M., and Rose, B.J. *Introduction to the Subjective Transfer Function Approach to Analyzing Systems.* Santa Monica, CA: Rand, 1984.

Watson, S.R., Weis, J.J., and Donnell, M.L., Fuzzy Decision Analysis. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. *SMC-9*, No. 1, January 1979, pp. 1-9.

White, C.C., A Posteriori Representation Based on Linear Inequality Description of A Priori and Conditional Probabilities. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-16*, 4, 1986, 570-73.

White, C.C., Sage, A.P., and Dozono, S., A Model of Multiattribute Decision Making and Trade-off Weight Determination Under Uncertainty. *IEEE Transactions in systems, Man, and Cybernetics, SMC-14*, 2, 1984, 223-29.

Winkler, R.L., Decision Modeling and Rational Choice: AHP and Utility Theory, *Management Science*, Vol. 36, No. 3, March 1990, 247-248.

Yager, R.R., "A New Methodology for Ordinal Multi-Objective Decisions Based on Fuzzy Sets", *Journal of the American Institute for Decision Sciences*, 1981, Vol. *12*, No. 4, pp. 589-600.

Zadeh, L.A., Fuzzy Sets as a Basis for a Theory of Possibility, *Fuzzy Sets and Systems*, Volume I, 1978.

Zeller, G.A., *A Standard Damage Assessment List (SDAL) for Lightly Armored Fighting Vehicles*. Final Report to Joint Munition Test and Evaluation Program Office, Eglin AFB, PA, ASI Systems International ASI89-09, Aberdeen, MD.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A

METHODS OF REPRESENTING AND ANALYZING CHOICE

A-1.0 BACKGROUND

A-1.1 GOAL

The work to be described attempts to integrate several very important aspects of the problem of choosing one military system from a potential set of available options. Modeling choice under uncertainty is a major part of the general problem of systems acquisition and also of the sub-problems of characterizing the potential costs and benefits associated with a potential system. Almost every problem involved in the description, effectiveness assessment, evaluation, and decision making with respect to major systems is also involved. This is also coupled with issues of organizational decision making and human cognition. The design of a robust, coherent, and usable procedure for choosing a system to fund (including research, development, acquisition) must address all of these issues and balance them appropriately, and this appendix attempts to clarify, organize, and make recommendations with respect to these issues. The particular focus of the work will be the analysis of the C3I warfare requirements.

The work interfaces with and draws upon work by SAIC to date with respect to evaluation using the Hierarchy of Objectives and Conditional Probability Logic as a means of evaluating options with respect to Naval Warfare requirements.

A-1.2 PREVIOUS WORK

A-1.2.1 Naval Warfare Requirements Definitions

Naval Warfare needs are described in a series of Top Level Warfare Requirement (TLWRs) documents. TLWRs have been developed by the Office of the Chief of Naval Operations (OP-07) for some Warfare Mission Areas (WMA), as well as for Electronic Warfare and for the Carrier Battle Force (CVBF). TLWRs are now being addressed at an even higher level, that of the Functions of the Navy, beginning with Sea Control and then for Power Projection.

The TLWRs for Sea Control are expressed in terms of Mission Success Criteria (MSCs). These are statements of objectives to be achieved in various mission situations. The ability to achieve the MSCs is expressed as a combination of Required Capabilities (RCs) in the various WMAs. The RCs are, in effect, sub-objectives, in terms of a possible combination of system capabilities, that would lead to the accomplishment of the MSCs. In the TLWRs for WMAs, these RCs become MSCs and, to support them, there is a set of RCs for platform mobility and sensor and weapon systems. C3I requirements have been stated subjectively in qualitative and quantitative terms in the TLWRs and other references, but not in a way that exhibits the contribution of C3I to Warfare goals.

Within the organization of the Deputy Chief of Naval Operations for Naval Warfare, the Director, Space and Electronic Warfare (OP-76) is responsible for the Space and Electronic Warfare (SEW) Appraisal, a major component of the Navy's Planning, Programming, and Budgeting System. OP-76 is also responsible for the administration of Team "C", which is guiding the development of the Navy's SEW Master Plan. This plan includes Command, Control, and Communications (C3) requirements. OP-76 is evolving a methodology for analyzing SEW requirements in support of these efforts. Previous work has resulted in the descriptive Command Process Model (NOSC Technical Document 1937), hereafter referred to as the CPM.

Within the Space and Naval Warfare Systems Command, the Naval Warfare Systems Engineering Directorate (SPAWAR-30) directs the development of architectural descriptions and assessments of current and future Naval Warfare Systems. The process is governed by the issuance of the

TLWRs by OP-07. In response, the Systems Engineering team is attempting to devise a means of providing a traceable accounting of the relationship between system performance and the TLWR. This has given rise to the development of a methodology for architectural description, modeling, and assessment that is ongoing. This methodology addresses Operational Functions, System Capabilities, and Force Performance Measures. This appendix provides support and guidance in coordination with that effort.

The objective of initial work was to develop a hierarchical multi-level analysis structure of functions and metrics, down to the Force level, that relates Operational Functions and Resource Capabilities to Mission Success Criteria, Required Capabilities, and Force Performance Measures, and describes how these depend on Mission context. The analysis structure makes evident the contribution of C3I, embedded in the operation, to effect Mission Success. The first report, entitled Hierarchy of Objectives (NOSC Technical Document 1938, Vol. 1), addressed an approach to functional analysis of Naval Warfare at the top levels, addressing military objectives and mission area characteristics to the intra-task force level, with a focus on how C2 affects results. The second report, entitled Decision Probability (NOSC Technical Document 1938, Vol. 2), presented methods for mathematically relating capabilities and objectives at those levels. That metric analysis is based on a common measure (conditional probability) to quantify the effect of dependency among functions at all levels of the hierarchy. The probability of making a decision affects what activities will take place, which, in turn determine what outcomes will occur. The system functions which support decision are discussed in the third vloume of this work (NOSC Technical Document 1938, Vol. 3) entitled Command and Control System Requirements Analysis.

A-1.2.2 <u>Overview of Approach to Warfare Requirements Characterization</u>

The following discussion is a summary of the approach described in NOSC Technical Document 1938, Volumes 1, 2, and 3.

A-1.2.2.1 <u>Role of Hierarchy of Objectives</u>. A Hierarchy of Objectives can be stated in terms of Missions, Functions, and Tasks. For a particular Force or System, its Functions are the activities it performs in order to accomplish its Mission. Its Tasks are its subfunctions, which are performed by its parts or subsystems. Mission objectives are based on achieving a preferred set of outcomes, which are particular states of the enemy's forces and ones' own, as well as the state of the environment, e.g., occupied territory. These objectives may support a higher objective, such as the capitulation of the enemy. The strategies, operations, tactics, and procedures used by each Force are a hierarchy of functions or processes that correspond with their hierarchy of objectives. The sub-objectives are to achieve favorable outcomes of the functions, i.e., those outcomes that contribute to achievement of the outcomes stated in the Mission objectives. Functions/objectives at each level may support several of those at a higher level or of a larger Force. The role of the Hierarchy of Objectives is to define the set of functions and their favorable and adverse outcomes, not only at the highest level, where the TLWRs establish Mission Success Criteria, but also for Required Capabilities and below. The sets of outcomes form the basis for defining a measure of potential achievement of the objectives.

A-1.2.2.2 <u>Role of Decision Making in the Hierarchy of Objectives</u>. Decision making is a function that is performed at all levels of the Hierarchy of Objectives. It is the function that determines which objectives to pursue, which functions to perform and which resources to use and when. The purpose of decision making is, therefore, to allocate resources to perform functions in support of higher objectives. The decision process consists of the performance of decision functions that involve interpreting information or choosing courses of action. These functions, called Command Functions, have decisions as their outcomes. They are described in the Command Process Model (NOSC Technical Document 1937 and Appendix A of NOSC Technical Document 1938, Vol. 1), which provides an updated version of the Command Functions previously used in OP-76 assessments.

A-1.2.3 Elements of System Description

A method of describing systems of all magnitudes was defined in the Hierarchy of Objectives report and is summarized in the Decision Probability report. The approach is similar to techniques in object-oriented programming. Objects are described by their states or attributes and by functions that are relationships between the states. Going beyond the basic ideas of object-oriented programming, this method recognizes that functions represent a causal conditioning between the states, i.e., that the value of a state (outcome of the function) depends on the values of other states. This conditional dependence can be the basis of a graphic technique for depicting architectural structure. The Hierarchy of Objectives is such a structure of functions to be performed in order to achieve preferred values of states. The conditioning relationship also provides a foundation for assigning a measure of likelihood to the values of a state that is dependent on the values of the other states. This can be a deterministic or stochastic likelihood that can be expressed as conditional probability or related measures.

A-1.2.4 Modeling an Architecture

Later in this appendix, the subject of modeling military worth using certain technical approaches in detail will be addressed. At this point, however, it is desirable to briefly discuss modeling of architectures. The concepts discussed underlie the approach taken in applying Conditional Probability Logic to the modeling of architectures.

In modeling an architecture, objectives can be stated as probabilities of attainment of preferred outcomes such as, the probability (confidence) that m units will survive a combat action. Capabilities of systems can be expressed as probabilities of outcomes characteristic of the functions of the system such as, probability of detection and of a false alarm. Modeling of tactics and procedures can be accomplished by examining appropriate conditional probabilities of higher level outcomes based on lower level events and capabilities as mapped out in the hierarchy of functions and objectives. A force architecture, then, is a combination of resources and associated system capabilities, procedures (tactics), and the decision making organization designed to achieve the objectives. A model of objective achievement conditional on scenario and system (force) architectures, which involves system capabilities along with force size, organization, and procedures, is a tool for assessing potential architecture performance.

A-1.3 RELATION TO ON GOING WORK

This brief characterization of the work to date describes a systematic approach to both characterizing Naval C3I requirements and developing a means of assessing potential performance with respect to meeting such requirements. The major question to be addressed by all this work is how to relate information about the potential effectiveness of proposed systems in achieving overall mission success to the question of choosing a development option. Note that a more specific question is how to relate meeting the required capabilities (RC) with respect to C3I to overall mission success. Indeed, answering this question with respect to C3I has been a major stumbling block for assessing the operational effectiveness of forces. C3I has been termed a force multiplier, and inherent in such a concept is that different capabilities with respect to weapons, sensors, and platforms combine in a complex, non-additive manner depending on how well the C3I works. Yet, many assessments of force capability treat C3I in an essentially subtractive fashion. That is, it is assumed that the weapons, sensors, etc. receive the C3I inputs they require in a timely, unambiguous form, and the systems are thus assumed to be supported by "ideal" C3I capability. Furthermore, decisions are implicitly modeled as activating the modeled procedure with probability 1.0 when the modeled situation occurs, without regard to whether the situation would be recognized by the decision maker or whether a particular procedure would be invoked by that decision maker. The synergistic nature of C3I capability (or the converse, the detrimental effect caused by the lack of it) is not systematically treated in most assessments of force capability. One simple reason is that it is extremely difficult to do so. Weapons and sensors can be simulated using highly detailed technical models. C3I involves command (decision probability and quality), control (procedures, and tools for implementing decision), communications, and intelligence. Many aspects of this involve humans and the associated questions of conditional behavior given different levels of automation in the organization. C3I can be simulated to a degree and the effects of C3I-Weapon-Sensor tradeoffs can be examined, but a disciplined process employing sophisticated tools is needed to do so. This issue will be discussed later in this appendix.

How much should be spent on C3I? When should it be spent? Should it be hardware, software, or training? How does all this relate to the same questions with respect to weapons and sensors and associated platforms? What are the tradeoffs? This somewhat wide ranging discussion emphasizes the complications involved in validly answering such questions. The rest of the appendix will attempt to characterize methods for attacking different parts of the problem as well as methods for integrating the different pieces. This will involve discussions of measurement theory, performance estimation, evaluation of worth, and organizational performance. The goal is a coherent statement of the current status of the problem areas and the approach to improved implementation of techniques for structuring elicitation and evaluation of judgments of system goodness and possibly automated tools for doing so.

A-2.0 THE PROBLEM OF CHOICE UNDER UNCERTAINTY

AS APPLIED TO SYSTEM EVALUATION

The general question of "How much hardware and associated software must be researched, developed, and acquired (RDA), and when?", is a good way to characterize the C3I system evaluation problem, for it is the practical problem to be solved. Too often, pieces of the problem are solved without interrelationships defined, resulting in the lack of a coherent methodology for answering the question. Two major questions must be initially answered. What is it that we wish to evaluate? What will be our metric for evaluation?

The question of what is to be evaluated is not simple. A DoD organization attempting to build a set of budget packages faces a different problem than a DoD team attempting to define the best Advanced Tactical Fighter (ATF) which in turn faces a different problem from a DoD organization trying to develop and evaluate a set of C3I requirements. In the budget case, options are defined as programs with associated work statements or plans, and are thus available for comparative judgments, although such judgments may involve complicated projections with respect to uncertain future benefit. The other two cases involve inferring a set of functional requirements with respect to a potentially vaguely specified future, and the difference between the two has mainly to do with system definition versus force definition. In both cases, however, it would be desirable to project the situation in which the force or system will have to perform and what the performance requirements will be. Because the issue is prediction with respect to an uncertain future, the issue of uncertainty is obviously relevant. Because the goal is to develop a capability to make sound consistent choices with respect to that uncertain future, choice as a decision under uncertainty is the second major issue. Thus the decision to implement a particular capability is discussed as a choice under uncertainty and is related to the large amount of work that has been addressed to that problem.

What would be ideal with respect to evaluating C3I requirements? First, an omniscient operational analyst-political scientist would project the future operational environment that C3I system capability would have to support. Representative scenarios that contained an accurately defined antagonist and associated situation would be defined. This is obviously a prediction problem commonly called threat assessment or some similar term. Next an operational analyst would indicate what performance levels forces with their associated platforms, weapons, sensors personnel, and C3I equipment would have to achieve as a function of this valid representation of the future. Typically, there are system tradeoffs here, and thus it would also be good to indicate ranges of tradeoffs between platforms, weapons, sensors, and C3I. Thus with a surveillance sensor of a given capability, the requirements for close in weapons capability to counter threat t_i is reduced. However, C3I systems must support the enhanced situation assessment capability associated with the enhanced surveillance sensor.

Already the problem sounds complicated, and it is; and it is also the case that a valid representation of the future, though an ideal, is rarely even approximated. Thus, a goal would be to examine a reasonably representative set of scenarios representing future operational requirements and from these deduce a set of desired military operations objectives (or associated MSCs) required. Further, from these deduce a set of C3I functional capabilities (or RCs) that would provide the support required to achieve the objectives (at least at some threshold level) in each scenario. Because the weapons and sensors are also hypothetical, this can be a difficult task.

Thus it is important, early on, to acknowledge that a critical part of this problem is about making a choice under uncertainty. What that choice problem is greatly determines the answers to many of the questions regarding such issues as to which level in the hierarchy of functions and objectives performance is to be modeled and at which level values and capabilities will be assigned and assessed, which types of assessments are required, the uniqueness and meaningfulness of results, etc. Consider two extremes. In one case, the problem has been so well defined that the actual forces to be involved in a combat scenario (one of the representative ones earlier chosen) can be described in detail, including battle group composition, capabilities of different platforms, locations of certain battle group functional capabilities, types and numbers of air squadrons on the battle group carriers, command and control capabilities such as surveillance, detection, acquisition, situation assessment, doctrine. All are so well defined that a conditional capability can be described parametrically for any feasible combination of relevant causal variables. In the second case, capabilities are described functionally and somewhat generally, possibly even qualitatively with terms such as good, fair, etc. Thus the overall ability of the force to engage threat t_i is good because the following capabilities exist, or a threat t_j can be detected at a range of 100 miles by sensor system S.

In the more detailed case, a fairly complete high-level regional engagement simulation could be supported by the information in the alternative definitions. In the latter, it obviously could not. In the former case, statements of the sort, "the probability of survival of platform P is .70 or higher until the number of incoming threats approaches n. At that point, given the scenario details, the probability drops rapidly and is very low as n approaches n_t". Such a global simulation is highly desirable for many reasons. Advocates of such simulations will indicate that warfare interactions among functional capabilities are often non-linear and even non-intuitive. Only in complex models can they be demonstrated. However, such models are prone to require many detailed constraints and assumptions that limit the generality of the results, and these limits are often not well understood or well communicated. Also, because such models are expensive to construct, their usage is subject to many practical constraints that can reduce the validity of their outputs. The more disaggregated functional approach described in the second case, on the other hand, can suffer such failures as those demonstrated by the following statement. "The battle group has a very good capability to engage air threats of type t up to numbers as great as n_i under the following conditions. The fleet also has extremely good electronic warfare capability thus enhancing the total anti-air function. Overall the described battle group gets very high marks with only a slight degraduation due to a near threshold anti-submarine warfare capability." This describes an evaluation using some sort of additive combination of high level functional requirements. Suppose that the fleet could be evaluated in the simulation and suppose further that it suffers rapid catastrophic defeat because the anti-air and EW capability are both resident on highly sophisticated carriers that are destroyed early by a sophisticated enemy submarine threat. Unless this problem is picked up in one of the functional requirements in the evaluation hierarchy, the evaluation could provide very misleading results with respect to performance uncertainty.

The example, though simple and well understood by practitioners of the art of system evaluation, illustrates that steps back from the omniscient prediction or the well-supported completely valid global simulation can lead to major errors. The answer to what is a best approach for a given problem of course depends on the specifics of the situation.

As indicated, one of the dependencies has to do with the degree of specification of the choice options. Another concerns the answer to the "bottom up" versus "top down" question. Many tools for choice depend on procedures developed from an axiomatic measurement theoretic approach. These may require assessment of preferences or similar comparative judgments, usually binary, with respect to observable options that form the bottom of a hierarchy constructed so that it starts at the top with global force performance and proceeds down to detailed comparable options. For example, the current Joint Decision Support System (JDSS) (version 2.2) has seven levels with the global level at the top of the hierarchy and specific capability objectives with respect to specific functional tasks seven levels down. A bottom up approach proceeds by building the higher level, more global assessments through aggregation of the lower level, more specific assessments.

Comparisons are often made with respect to expected benefit. Top down approaches proceed by addressing the relative importances of the high level objectives and proceeding down by assessing potential relative contributions of lower level objectives to the higher ones. Such top down judgments, through possible, must be accomplished very carefully with clear definition of what in fact is being assessed. The concepts of benefit and importance are often not well-defined (see Section A-4.1 to A-4.3 for further discussion) and may relate ambiguously to different objectives at different levels of the hierarchy. Numerical scales for these value judgments do not in themselves make valid tools for evaluation unless the scales are calibrated to the objective in question. Often these are problems with hierarchical non-probabilistic, non value-based approaches.

In this report, we discuss many methods and highlight pros and cons of these with respect to the evaluation of C3I requirements. There are many potential approaches to such evaluations, and although many are not necessarily wrong, they are not appropriate for the defined constraints both theoretical and practical, and this will be further discussed. A method must systematically accommodate necessary steps back from the ideal evaluation. Such steps are motivated by very obvious practical constraints such as the following:

- the inability to validly characterize the environments and associated enemy capability in which the battle force must operate
- the inability to develop valid models to address all the complex functional dependencies involved
- the problems humans have with respect to certain assessments of uncertainty and worth
- conflicting organizational objectives for the evaluation.

A criterion for any approach developed would be that it employs assessment techniques consistent with a correct measurement theoretic approach. A second is that the assessments required are compatible with the known results of the vast literature on human decision making and cognition that clearly delineates certain limits and biases that are common across individuals. A third is that the procedure is implementable from an organizational point of view. Given these as general goals, the following sections discuss some of the different approaches.

A-3.0 OVERVIEW OF METHODS FOR PRESCRIBING DECISIONS

There are numerous approaches to prescribing decisions involving uncertainty and value tradeoffs. One of these is the well-known decision analytic paradigm based on a combination of expected utility theory and the Bayesian statistical approach to uncertainty assessment. We shall call this Decision Analysis in the formal or narrow sense. Other methods involve alternative approaches to uncertainty assessments (e.g., possibility theory), methods for ordinal as opposed to cardinal assessments of uncertainty and value, alternate methods for structuring and scaling hierarchical value functions (e.g, the Analytical Hierarchy Process - AHP), and procedures for allowing imprecision in cardinal utility or probability inputs (e.g, Imprecisely Specified Multi-Attribute Utility Theory, ISMAUT) and Fuzzy Decision Analysis. All thes· methods are part of a broader addressal of the problem of decisions or choice under uncertai:·y having some potential advantages over formal Decision Analysis and also having some shortcomings. This section of the appendix will consider several of these approaches to addressing decisions and evaluations. The first will be traditional Decision Analysis.

A-3.1 OVERVIEW OF FORMAL DECISION ANALYSIS

Decision Analysis is a method for addressing decisions that involve uncertainty and value tradeoffs. Decision Analysis can be defined in the narrow sense as based upon the axioms of expected utility theory, in particular that a decision maker acts (or should act) by choosing the alternative, A, which maximizes the expected utility:

$$EU(A) = \int u(x) \bullet p(x \mid A)\, dx \qquad \text{(A-1)}$$

where x is an outcome in the domain, X, of attributes of the outcome, $p(x \mid A)$ is the probability of the outcomes given a choice of A, and $u(x)$ is the utility of each attribute outcome.

The approach characterizes a decision as an initial choice point, that is followed by a sequence of subsequent acts and events (See Figure A-1). Each subsequent act is conditional on the sequence of acts and events up to that point and is a choice point. Note that acts are actually the implementation of a decision to act in a certain manner. In Decision Analysis, the decision to act and the implementation are characterized in the tree simply as acts. Each subsequent event can have several potential outcomes (and can, in fact, best be characterized as a random variable). The structuring of the set of potential acts and events is part of the decision analysis method that yields a decision tree which contains a large number of branches, each branch representing one of the many potential act-event-outcome combinations. Eventually the decision tree branches terminate in an alternate set of event outcomes having consequences that are represented as levels on attributes that characterize the values at stake in the decision. Typical military decision attributes would be operational performance, casualties, supportability, acquisition cost, etc. A utility function can be assessed over these attributes, thus assigning a worth to an act-event-attribute level combination. The procedures for assigning utilities to multi-attributed alternatives have been variously labeled MAUT, MAU, MAUA, etc. (See Keeney, R.L., and Raiffa, H., 1976, for an exposition of Multi-Attribute Utility Theory.) MAUT will be used in this report.

The Decision Analytic paradigm is very important to the overall thrust of this report and will be used as an appropriate foundation with which to compare other approaches to evaluation of systems. In other words, the evaluation of C3I capability is treated as a choice among alternatives, which result in the uncertain outcomes of operational objectives and developmental and maintenance costs, or other attributes. Decision Analysis is one prescriptive approach for such decisions which will, in theory, provide a normatively correct analysis of the decision problem. However, as a practical solution, such an analysis is difficult to envision for many reasons, one of which is the sheer magnitude of the uncertainty representation. As discussed in the previous section, compromises must be addressed with respect to such issues as validly representing the uncertainty, problems with respect to human assessment capability, etc. Suppose for now, however, that we could evaluate C3I systems using the Decision Analytic approach. How would we proceed?

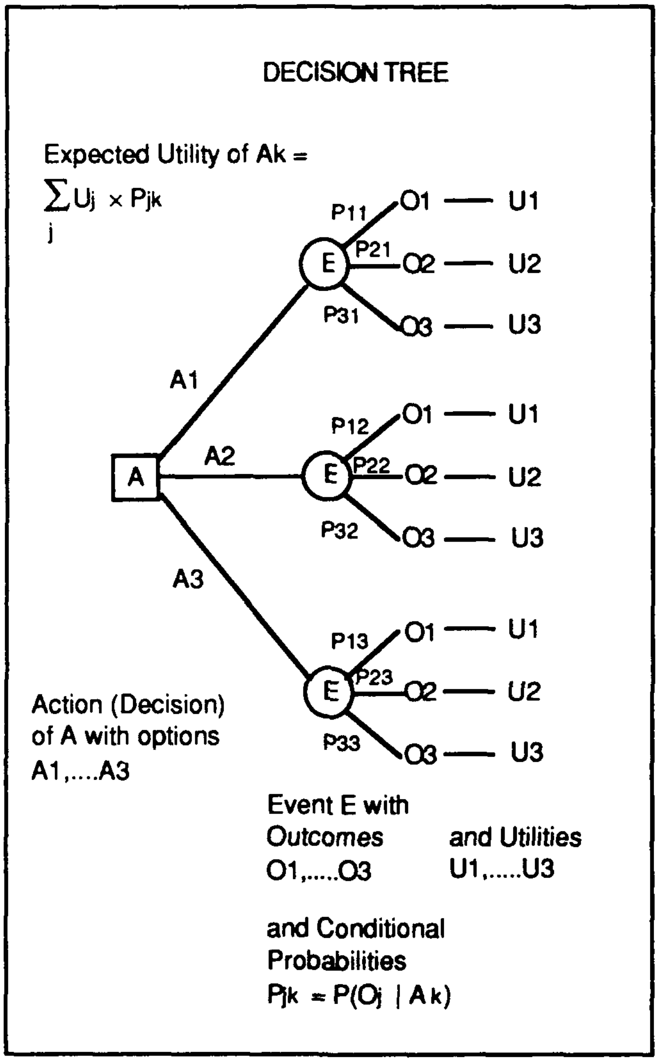


Figure A-1. Decision Tree for One Act and One Event

First we would define our decision alternatives, uncertain events, potential consequences, and relevant sources of information. What is the choice problem? Is it to choose one of three C3I global capabilities? Is it to find some combination of R&D, system upgrades, current systems, training, etc? Whatever it is (and this does make a difference to the issue of uncertainty and value in the decision), the alternatives that span the set of available choices are defined. We also characterize the decision makers value and uncertainties clearly. These will later be refined with greater precision, but we need to know what is at stake and what are the key uncertainties. The next sections discuss some of the major issues with respect to uncertainty and value.

A-3.2 CONDITIONAL DEPENDENCIES, INFLUENCE DIAGRAMS, AND COALESCENCE

Given clear definitions of the decision, alternatives, uncertain variables, and values involved, the next step involves developing a structural representation of these variables. This actually involves several steps. One is to lay out the decision tree as the appropriate sequence of acts, events, and consequent attribute levels. Because the issues of conditional independence among events are quite complex, laying out the tree can often be assisted by using an influence diagram to structure the events and their conditional interdependencies. In such an influence diagram, events are represented as nodes, and conditional dependencies as arcs between nodes (See Figure A-2). Such a representation helps to represent the complex event interdependencies that can be difficult if only using a decision tree (See Call, H.J. and Miller, William A., 1990, for a detailed discussion of some of these issues. See also Shachter, R.D., 1986). The decision tree provides a clear representation of the dependencies between acts and events, which cannot be well done with influence diagrams, and thus the integration of influence diagrams with decision trees provides for an efficient representation of the act-event chain, and the usually very large tree associated with such a structure.

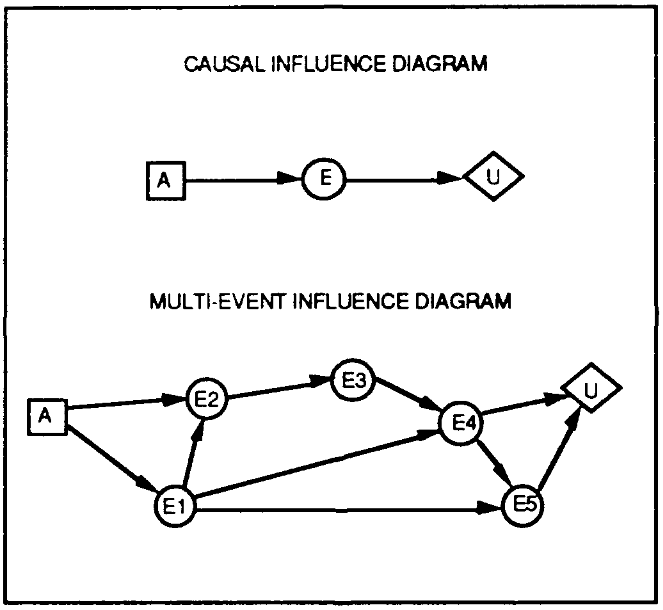


Figure A-2. Influence Diagrams (Causality Net Form)

Decision trees typically become quite large for even small problems. Simple mathematics will convince the reader of the combinatorial issues involved. Thus it is desirable to reduce the size and detail in a problem by reducing the size of the tree structure. One way is by addressing conditional dependencies. Coalescence is the process of collapsing a tree by eliminating certain sub-trees in the general tree structure that can be characterized by certain probabilistic independence conditions among the variables. For example, suppose we have three binary events E_1, E_2, and E_3 with outcomes O_1 or $\neg O_1$, O_2 or $\neg O_2$, and O_3 or $\neg O_3$, respectively. This implies eight branches in the decision tree. Suppose further that the value function is assessed directly as a function of the event combination (single attribute) and is independent of the outcome O_1 or $\neg O_1$ of event E_1. Then the value need only be assessed four times and the value V over the outcome vector (O_1, O_2, O_3), $V(O_1, O_2, O_3)$ can be represented as $V(O_2, O_3)$ with appropriate probabilistic weights derived by coalescing over O_1. (This issue will arise again in discussion of the conditional probability logic in which events are represented as variables and coalescence involves integrating over the appropriate random variables according to the conditional relation specified by the mathematical representation of the structure.) Implementation of influence diagrams is similar to implementing graphical techniques in general. That is the algorithm employed does not compute all possible branches. Rather the conditional dependencies are examined for opportunities to coalesce and thus eliminate nodes from the structure. This capability is provided through the direct statement of conditional independence indicated in such diagrams.

At this point to prevent further cumbersome notation, it should be noted that events and acts, though different conceptually, can be treated mathematically as random variables. Subsequent acts can in fact be considered to be events in a more general representation of the Decision Analytic paradigms. For purpose of brevity and efficient discussion, events, event outcomes, and subsequent acts will be denoted by the general term variables such as X, Y, and Z, each of which can be a multi-dimensional vector (e.g, X could be an event having potential outcomes ($X_1, X_2, \ldots, X_m$); Y could be an act having options ($Y_1, Y_2, \ldots, Y_n$); etc.)

The uncertainties involved in the decision are further addressed by conducting sensitivity analyses which often point to a need for a revision of the structure. Also, potential information to be obtained in order to reduce uncertainty is examined to assess the cost/benefit issue of the additional information, i.e., is the added expected utility worth the sure loss of the cost of the information? If so, the information can be purchased and the decision tree revised. Such sensitivity analyses help to assure the integrity of the representation of the problem.

A-3.3 MEASUREMENT THEORY FOUNDATIONS

The structuring of the decision problem using decision trees and influence diagrams provides the foundation for the assessment of probabilities and values. Decision Analysis uses probabilistic assessments based on the axioms of probability theory. Similarly, the multi-attributed value function that is assessed over the multi-attributed outcomes is based on the axioms of expected utility theory. That is, Decision Analysis is an implementation based on a measurement theoretic foundation of expected utility theory. This will be discussed in more detail later in order to motivate comparisons with other approaches. However at this point, it must be clear that probabilities are defined on a 0-1 scale and the probability of the union of two events will, in general, be greater than the probability of either one if they have a finite intersection (from a set theoretic point of view) and one doesn't contain the other. This is to differentiate probability from certain other uncertainty measures such as the possibility measure of fuzzy sets (See Zadeh, 1978). Similarly, utilities are cardinal utilities measured on an interval scale, thus unique up to a linear transformation. The measurement theory approach provides a representation theorem that states necessary and sufficient conditions for a correspondence to exist between an empirical relational system - in this case the set of decision alternatives with an assessment defined over pairs of alternatives, say a preference - and a numerical relational system - in this case the real numbers with the usual full order that provides for multiplication and division. The axioms of expected utility theory provide a set of tests for this empirical relational system including transitivity as well as certain independence tests, e.g., that utilities are independent of probabilities. Satisfaction of the tests supports the theory that the decision maker providing the assessments maximizes his expected utility in making decisions. Thus his assessments can be used to construct utility and probability functions and to make prescriptive recommendations. This makes a very rigorous, tedious, process sound simple, and it is obviously not, but the important point is that the measurement theory provides a theoretical foundation for the assessments and the resultant numerical representation. This point is emphasized here because certain other approaches, though axiomatically based, do not provide such a clear linkage between the theory and the required assessments. The Analytical Hierarchy Process is an example and will be discussed later (See Saaty, T.L., 1980).

A-3.4 MULTI-ATTRIBUTE UTILITY THEORY (MAUT)

As indicated in section A-3.0 and Figures A-1 and A-2, the decision tree and the structuring of acts and events intervening between the decision and the outcomes provide the means of attaching a probability to the utility associated with each potential outcome. However, each outcome can be characterized by levels on a potentially large number of attributes $X_1 \ldots X_S$. In order to assess the expected utility, it is necessary to be able to attach a value V or utility U to each combination of attribute levels. That is if terminal outcome O_{iT} is characterized by the set of attribute levels (x_{iT1}, x_{iT2}, x_{iTS}), we need to asses $U(x_{iT1}, \ldots x_{iTS})$.

A-3.4.1 The MAUT Theory

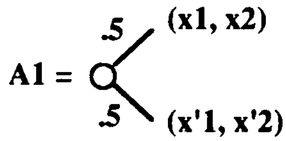
The particular form of the multi-attribute utility function is developed using the procedures of multi-attribute utility theory (MAUT). The appropriate form of the function $U(x_1, \ldots, x_S)$ is determined by using a series of value-wise independence tests (See Keeney, R.L. and Raiffa, H. for a detailed discussion of MAUT). Recall that the Utility function U is the result of using MAUT to test the axioms of the expected utility theory. The axioms deal with preferential independence. For purposes of our discussion, consider three attributes of alternatives $X_i = (x_{i1}, x_{i2}, x_{i3})$. Attributes X_{i1} and X_{i2} are considered value-wise independent if preferences among X_i, X_j, where X_{*1} and X_{*2} levels vary but X_{*3} level remains fixed, are independent of the fixed level of X_{*3}.

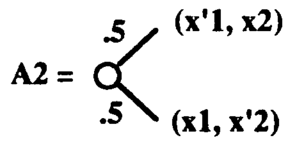
Another even stronger condition known as utility independence involves risk and has the following characteristic. X_1 is utility independent of X_2 if conditional preferences for lotteries on X_1 given X_2 do not depend on the particular level x_2 of X_2. Further define attributes $X_1 \ldots X_S$ as additive independent if preferences over lotteries on $X_1 \ldots X_S$ depend only on their marginal probability distributions and not their joint probability distributions. If it is true that for all attributes $X_1, X_2, \ldots, X_S$, the X_i are additive independent then we have

$$u(x) = u(x_1, x_2, \ldots, x_n) = \sum_i k_i\, u_i(x_i), \qquad \text{(A-2)}$$

When total additive independence fails, the utility function takes on a more complicated form depending on the nature of the failure of independence.

A simple two attribute example displays additive independence. Consider the following two alternatives A_1 and A_2.





Additive independence implies that the decision maker will be indifferent between A_1 and A_2 because the marginal distribution over the levels of $X_1 = (x_1, x'_1)$ and $X_2 = (x_2, x'_2)$ are the same for both alternatives even though the joint distributions are not. The example generalizes easily to the S dimensional case. Thus $X_1 \ldots X_S$ are mutually preferentially independent if preferences among lotteries involving the attributes depend only on the marginal probability distributions over the attribute levels for each alternative. Alternatives with the same marginal distribution over attribute levels are equally preferred.

If the decision maker's preferences among such lotteries do not satisfy the additive independence condition but satisfy utility independence [lotteries over X_i do not depend on the complement of X_i, the levels at which $(X_1, X_2, \ldots, X_{i-1}, \ldots, X_{i+1}, \ldots X_S)$ are set] then we have the multiplicative form of U which is stated as follows:

$$k\ u(x_1, x_2, \ldots, x_n) + 1 = \prod_i [\ k\ k_i\ u_i(x_i) + 1], \tag{A-3}$$

If both additive independence and utility independence or preferential independence fail, the form of U is more complicated. (For such a discussion, see Keeney, R.L. and Raiffa, H., 1976.)

Note that the reason for desiring additive independence, which is a major assumption subject to test, is to simplify the calculation of expected utility. The additive independence case collapses the decision tree by assessing alternatives on the basis of their expectations in terms of utility on attributes separately. Remember that each branch of the decision tree is but one potential pathway to the multi-attributed outcome for each alternative, and different paths involve different joint outcomes over attribute levels. Thus this condition is required and should be tested for additive expected utility theory to be employed.

A-3.4.2 Scaling the Utilities in MAUT

Suppose we are choosing among alternatives $X_1 \ldots X_m$ characterized by levels on attributes $X_1 \ldots X_S$. Thus $X_j = (X_{j1}, X_{j2}, \ldots, X_{jS})$. Further define for each x_j two levels x_{jo} and x_{jv} such that the following is true. Fix all levels of $X_i \neq X_j$ at nominal levels x_{iF}.

Define $U(x_{1F}, x_{2F}, \ldots, x_{jo}, \ldots, x_{SF})$ = minimum of $U(x_{1F}, x_{2F}, \ldots, x_j, \ldots, x_{SF})$ for all x_j. Further $U(x_{1F}, x_{2F}, \ldots, x_{jv}, \ldots, x_{SF})$ = maximum of $U(x_{1F}, x_{2F}, \ldots, x_j, \ldots, x_{SF})$ for all x_j.

Then $$W_j = U(x_{1F}, x_{2F}, \ldots, x_{jv}, x_{SF}) - U(x_{1F}, x_{2F}, \ldots, x_{jo}, \ldots, x_{SF}) / K, \tag{A-4}$$

where $$K = \sum_j [\ U(x_{1F}, x_{2F}, \ldots, x_{jv}, x_{SF}) - U(x_{1F}, x_{2F}, \ldots, x_{jo}, \ldots, x_{SF})\ ]$$

With this definition, we can set $u_j(x_{jo}) = 0$ and $u_j(x_{jv}) = 100$ and the other x_j can be scaled relative to these. The W_j become rescaling constants. The procedure for assessing the W_j can be deduced from the equations for the weights. Note however, that the equations for the weights involve hypothetical alternatives.

The values $(x_{1F}, x_{2F}, \ldots, x_{jo}, \ldots, x_{SF})$ and $(x_{1F}, x_{2F}, \ldots, x_{jv}, \ldots, x_{SF})$ are not necessarily among the attribute outcomes, that may actually occur, associated with any of the choice alternatives. This is not a problem, for the choices requested are probably quite meaningful. That is, the assessor is asked to make a preference between two well-defined, meaningful outcomes even if they are not among the actual ones that result from the options being considered.

Thus the procedure should be clear. Once the tests for independence have been satisfied verifying the existence of an additive representation, then the intra-attribute utility functions are established by fixing the other attributes at meaningful levels and varying the levels of one. The inter-attribute weights can then be established by comparison of inter-attribute swings from the minimum to maximum on one attribute as compared to the same swing on another, or potentially several others.

A-3.4.3 Value Independence versus Stochastic Independence

Two kinds of independence have been discussed in the report, independence in a stochastic process and independence of preferences and thus values. Put simply, two events are stochastically independent if the probability of their joint occurrence is the product of their respective marginal probabilities. Two attributes are value-wise independent if preferences among options varying with respect to one attribute are independent of the other attribute. Stochastic independence refers to random variables. Value-wise independence has to do with preferences over attribute levels. In choosing jobs, preferences over the type of work I might do and the quality of my commute to work can be value independent even though the two are highly correlated in likelihood. Certain jobs can only be obtained in large metropolitan areas with which are associated nasty commutes. Similarly, I cannot assess my value for differing levels of police force quality independently of where I live (and thus my need), but the two could be probabilistically independent. Thus the two types of independence are different.

A-3.5 SUBJECTIVE EXPECTED UTILITY

Decision Analysis has been described as a prescriptive theory, indicating that given a few testable independence conditions, utility functions can be derived and expected utilities of alternative options calculated. Prescriptive decision theory recommends that people behave as a rational, economic people. That is they should be rational and they should maximize something. In the Decision Analysis case, the something is expected utility. For a period of about twenty-five years expected utility theory was tested as a descriptive theory of people's choice behavior and found to be lacking mainly due to people's unfamiliarization with the type of decisions offered. Given failure of expected utility, a more general theory is subjectively expected utility (SEU) in which people are assumed to assign subjective probabilities to events and utilities to attribute levels and to combine the two appropriately using the mathematics of utility theory. Because people were often presented two outcome gambles, and SEU theory provides for two transformations, the theory was well suited to modeling the task and not necessarily actual behavior. As experiments became more sophisticated, SEU was found wanting as a descriptive theory. A major question involves its appropriateness as a prescriptive theory. It is often assailed for recommending that a decision maker faced with a one time choice should choose on the basis of an expectation (which in fact usually is not one of the potential outcomes). Why not minimize expected loss or treat the whole problem as a game against (a possibly antagonistic) nature? Or why not present the decision maker with the range of uncertainty of the utility of the outcomes?

These questions will be left aside, but an important issue is the transformation of probability. It was found in descriptive studies that people's probabilities do not obey the axioms of probability theory, do not add to one for complementary events, etc. Clearly, many processes do not have well-defined "objective" probabilities in that they involve future events with a vaguely defined event generating process. Subjective probabilities are thus required, but the fact that people's behavior may be different for judgmental assessments of uncertainty where there is no clearly defined random variable and associated event generating process leads to the consideration of methods of assessment that may be more compatible with people's cognitive processes. Such compatibility is, of course, a new criterion for evaluating an approach.

A-3.6 THE GOODNESS FIT OF THE FORMAL DECISION ANALYSIS APPROACH

As we discuss goodness of fit from a statistical point of view, we ask how well the data conform to a postulated theoretical form. Is the departure within the bounds of error given the postulated theory or not? If not, we can discard the theory barring some clear reasons for the lack of fit. The Decision Analytic paradigm is a very solid one for characterizing the process of choosing among defined alternatives. Alternatives are defined. The stage is set by defining valid scenarios. The scenario details are generated by examining the scenario detail - alternative interactions. Terminal event outcomes are projected. Levels on attributes are assessed for the outcomes. Values are assessed over attribute levels. Appropriate checks with respect to stochastic and value-wise independence provide for tests if the structural fit of the theory. Influence diagrams, MAUT tools, and decision trees are well defined for developing conditional probability and value assessments thus reducing the potential for error.

In order to assess "goodness of fit" of an approach to evaluation, the issue of what is error must be addressed. In Decision Analysis, errors of fit can be due to violations of different kinds of independence. The use of influence diagrams and decision trees helps to ensure that the structure is appropriate. Violations of independence may lead to re-structuring, because the structure of the decision tree is ultimately a function of the empirical relational system, the preference structure and the probability estimations.

What if the alternatives are not well defined? How would a Decision Analysis proceed? Indeed, if there are no alternatives, is there a decision to be made? This becomes a major question for certain "top down" approaches. Decision Analysis has a normatively correct foundation and provides a prescriptive recommendation based on that foundation. Further, assessment procedures are based on that foundation and a basis for statements about errors can thus be established. These characteristics are criteria by which other approaches can be compared and evaluated.

- Clarity of Purpose Is the decision or evaluation problem defined? Is it clear what the goal of the analysis is to be?

- Theoretical Coherence Is there a theoretical basis for the approach? Is the approach an axiomatic one consistent with or translatable to some version of measurement theory?

- Procedural Coherence Are measurement procedures clear? Are they derived from the theory?

- Error Theory Is there a way to assess the goodness of fit of the structure? Is there an error theory?

- Usability - Practicability Does the approach satisfy the real world constraints of simplicity, time consumption, organizational fit, implementability, repeatability, and explainability?

If we consider the formal Decision Analytic approach with its supporting software and methods developed, it scores well on clarity of purpose, theoretical coherence, and procedural coherence. There is a problem with errors, for the methods properly implemented use preferences and probability assessments and these are, indeed, subject to unreliability and error. When is an error simply unreliability or imprecision, and when is it in fact a violation of the assumptions? A method for specifying expected levels of error is desirable and the well-structured framework provides a foundation for such an error theory. One possible move in that direction for utilities at least, is ISMAUT discussed in Section A-3.7.2.2.

It is worth mentioning somewhat as an aside, albeit a very important one, that in practice, such measurement theoretically based tests for additivity are tedious and time consuming, and the inevitable tradeoffs between theoretical correctness, precision and accuracy, and practical constraints enter the picture. For obvious reasons, the practical constraints often rule. This issue is a major one for this report that must be given very deep consideration. Too often practioners of the trade and proponents of various approaches battle each other from different vantage points. One argues theoretical validity, another total assessment accuracy, and still a third overall usefulness. None are necessarily correct, but it would be good if these issues were addressed separately, clearly, and coherently.

On Usability - Practicability, the evaluation of Decision Analysis is mixed. A problem even moderately complex can tax the methods causing the decision tree to expand extremely rapidly and thus demanding rigorous attention to coalescence and simplification. For example, a scenario containing nine events, each having three outcomes, all of which influence the value function to require a tree with nearly 20,000 branches! (Most scenarios are much more complex!) Further, to conduct all the probability and value independence tests is a practical problem, for the decision maker typically will not sit still for such a voluminous number of required assessments. This has been a problem applying Decision Analysis. Tailored software and the use of trained decision analysts helps move the process along, but the procedure must be evaluated as having problems on this criterion.

A-3.7 METHODS UNDER THE BROADER DECISION ANALYSIS DEFINITION

A-3.7.1 Methods Providing for Uncertainty Assessment

A-3.7.1.1 Imprecise Probabilities. One of the problems with Decision Analysis already discussed is the assessment of probabilities. In a measure theory approach, the conditional probability of A given B, $P(A \mid B)$, is the measure of the set $A \cap B$ divided by the measure of B. Thus .

$$P(A \mid B) = P(A \cap B) / P(B). \tag{A-5}$$

The axioms of probability are defined so that the P-measures of complementary events sums to 1.0. Further, the use of assessment devices such as the well-known probability wheel used by certain practitioners of Decision Analysis (in which a circle can be divided into two colored sections much as a pie is sliced into two parts) helps to ensure such coherence. However, a major result of applied work is that experts have problems with probability assessments. Part of the problem concerns the definition of probabilities. Bayesians define probabilities in terms of behavior. Thus if I am indifferent between \$X for sure (where X is a number between 0 and 1.0) and a gamble in which I receive \$0 if an event does not occur and \$1 if A occurs, then the subjective probability of the event A can be judged to be X (see DeFinetti, 1951 for further details). The problem with this definition is that it is hard to use in practice. Further, it involves risky decision making which is well known to have a host of problems associated with it.

Other definitions of probability involve issues of degrees of belief, estimation, symmetry on a sample space, etc. The literature is too voluminous to discuss here, but a definite problem in assessing probabilities is the lack of well-defined measurement procedures such as those used for utilities. (This does not mean that such procedures do not exist. Rather, the meaning of the questions that the person must answer are often much less clear than the preferences called for in utility assessment procedures.)

Two general approaches have been developed to relax the rigor with which probabilities are estimated, neither of which solves the above problem, but both of which address some practical issues.

The first is fuzzy sets as described by Zadeh (See Zadeh, 1978) in which the membership of elements in sets is not a zero-one function. Rather an element can have a degree of membership in a set. Thus, the degree to which John is tall can be expressed by a degree of John's "tallness", the degree to which he belongs to the set of tall people. This definition led to the concept of a set with imprecise boundaries and thus fuzzy probabilities or "possibilities". The literature is voluminous with respect to the usefulness and even correctness of this theory and the arguments will not be repeated here. One issue is important, however, and that involves Zadeh's definition of the operations of set intersection and set union. Thus with fuzzy sets,

$$\text{Poss}(A \wedge B) = \text{Min}[\text{Poss}(A), \text{Poss}(B)], \text{ and} \tag{A-6}$$

$$\text{Poss}(A \vee B) = \text{Max}[\text{Poss}(A), \text{Poss}(B)]. \tag{A-7}$$

The use of such definitions for probability would certainly cause problems with respect to the assessments of conditional probabilities in decision trees, and thus the use of such possibilities would seem to be problematic from a Decision Analytic view point. (However, see Watson et. al., 1979, for a different viewpoint and detailed discussion. See also Freeling, 1980.)

Another approach to the probability problem is to treat a probability of a binary event as a random variable. Thus, the probability, $p(x_i)$,that a random variable X takes on some value x_i is itself a random variable, for which the measure is $p(p(x_i))$. This allows the assessor to provide several parameters to describe the probability. He can describe the maximum likelihood estimate of the probability as well as some range, even a confidence interval. Although this approach causes problems for some purists, from a practical point of view, it could be incorporated quite easily. Thus the probability of a single event A could be entered as a distribution. This becomes a bit more sticky when one must assess probability distributions for continuous random variables, e.g., energy supply in the year 1991, etc. These are already assessed as distributions over the random variable. Allowing the imprecision would involve second order distributions. Again, it is possible but is less easy to grasp than the imprecise assessment for a single discrete event.

Nonetheless, there is a practical problem that probabilities, like values, are subject to imprecision, especially if more than one expert is asked about the event or variable in questions. One way to solve the problem routinely used in practice is to conduct sensitivity analyses to assess the impact of such imprecision in the decision structure. Often there is little effect. This works fine unless there is a similar problem with a large number of variables, in which case sensitivity analyses become a practical problem from a time and complexity viewpoint.

A-3.7.1.2 Some Applications Involving Imprecise Uncertainty Measures. In response to apparent problems of decision makers in assessing probabilities, and the work by Zadeh (1978) on possibility as a substitute for probability, attempts have been made to relax the need for probabilities as specifically defined by the conventional axioms of probability theory. This section will briefly discuss some of that work. There is a great deal of confusion about exactly what imprecise probabilities are. In order to have a foundation from which to consider this confusion, the measurement theoretic viewpoint will be utilized. The process of establishing a representation theorem for probabilities consists of establishing the correspondence between an empirical relational system and a numerical relational system. Depending on the properties of the empirical relational system, say the events, and the assessor's judgments with respect to relative likelihoods, the resultant event probabilities will be measured on a scale unique up to some transformation. If the probabilities obey the usual laws of probabilities, e.g., complementary event probabilities sum to 1.0, the logical rules for "and" and "or" with respect to events, etc., then the permissible transformation on probabilities include only the identity transformation, that is, the probabilities are fixed in the 0-1 real number interval.

The fact that probabilities are precise does not mean that the assessment must involve a cardinal judgment. Rather, cardinal probabilities can be derived from preferences among lotteries which corresponds to the Bayesian axiomatic treatment of subjective probability (See DeFinetti B., 1951). Thus, the empirical judgment required and the empirical relation can be an ordering of event likelihoods. However, certain independence and indifference axioms must be satisfied. Thus, in discussing imprecise probabilities, the nature of the imprecision must be clarified. For possibilities, the imprecision is in set membership and thus the axioms pertaining to set unions and intersections are different.

Other approaches discuss imprecise inputs such as event orderings or numerical ranges on event likelihoods. Note that event orderings in fact provide ranges on event likelihoods. The issue becomes the desired uniqueness of the resultant uncertainty measures. For example, certain approaches discuss a probability as a random variable. In that case, input assessments would have to provide sufficient information to yield a distribution of some specified form. Thus the question with respect to precision is, in essence, what is the uniqueness of the scaled numerical probabilities derived from the input assessments and resultant empirical relational system and what are the properties of the empirical relational system, i.e., what are the required inputs and what conditions must they obey?

A-3.7.1.3 <u>Fuzzy Decision Analysis</u>. With the advent of possibility theory and fuzzy sets, decision analysts investigated the use of imprecise uncertainties in Decision Analysis. Watson, Weiss, and Donnell (1979), investigated the use of fuzzy logic and concluded that Decision Analytic theory could be modified to allow for such imprecise inputs. Using fuzzy probabilities, the input uncertainty assessments were of the form, "The degree to which X belongs both to A and to B is equal to the smaller of the individual degrees of membership" (p. 3). The output of the assessment process for the expected utility of a gamble involving choosing alternative A with potential outcomes A_1, with possibility p, and A_2 with possibility (1-p) would yield the expected utility of choosing A : $\mu_A(a)$, for $a = pu_{A_1} + (1-p)u_{A_2}$. The corresponding output for Fuzzy Decision Analysis is

$$\mu_A(a) = \max_{pu_{A_1} + (1-p)u_{A_2} = a} [\min\{\mu_P(p), \mu_{A_1}(u_{A_1}), \mu_{A_2}(u_{A_2})\}] \qquad \text{(A-8)}$$

where $\mu_X(x)$ is the degree to which x belongs to the possible sets of expected utilities for gamble X and $\mu_P(p)$ is the degree to which p belongs to the set of possible values for the probability in the defined gamble involving A_1 and A_2. Needless to say, such discussion is difficult to follow for one unaccustomed to using fuzzy probabilities and utilities. (The reader should consult Watson et. al. for detailed discussion of these concepts.)

Freeling (1980) investigated both fuzzy probabilities and possibilities. He concluded that possibilities required an operational definition in order to justify using them in place of standard Decision Analysis probabilities. Freeling concluded that fuzzy probabilistic decision making is a "consistent prescriptive device". He further discusses the coherence of a decision maker as perfect integration of all knowledge relevant to the uncertainties at hand resulting in the ability to provide completely consistent responses concerning probability assessments. He shows that the fuzzy extension of probability theory is not inconsistent with the traditional Bayesian approach, but rather is an extension of that approach. Further, he feels that the approach shows promise for helping with the reconciliation of inconsistent probabilities.

Another application with respect to imprecise uncertainty assessments is a tool described in detail in Larsen and Dillard (1989). They report on a fuzzy logic decision tool developed at NOSC for solving multiple objective decision problems where the objectives have differing degrees of importance. The algorithm is claimed to have relevance to command and control problems similar to those discussed in this effort where the situation is complex and objectives are hierarchical and possibly vague. Inputs to the algorithm include a set of choices, a set of objectives, the degree to which choices satisfy objectives, and a comparison of the importances of objectives. Two methods are discussed for comparing the importance of objectives. One is a rank order method proposed by Yager (1981) based on fuzzy sets. The second is based on Saaty's (1977) pairwise comparison process. The resultant fuzzy logic tool has been implemented and used to aid in expert system development planning as well as in air strike planning.

These are only a few of a large number of approaches to alternatives to the conventional implementation of probabilistic assessments. Such efforts are bolstered by several empirical issues. One is that people have cognitive limitations with respect to processing requirements in assessments of uncertainty. (See Kahneman, D., Slovic, P., and Tversky, A., 1982.) Also, experts have difficulty with respect to cognitive models of uncertain environments, and thus the basis for assessment of conditional probabilities must be suspect. Third, even if experts have reasonable cognitive models, precise judgments with respect to uncertainty are far more difficult than say, preferences for options in utility assessments.

Thus, procedures for assessing probabilities or uncertainties must address these difficulties. Note that the solutions are not the same. Allowing imprecise probabilistic inputs will not solve the problem of inappropriate or inadequate cognitive models and, in fact, will likely only make it worse. Using fuzzy set inputs, where set membership is actually crisp, is not desirable. Any solution must therefore first question what the source of the problem is. If it is in the structure, perhaps more use of influence diagrams and models is required. If it is in the expert's thought process with respect to uncertainty, perhaps training or reformulating the assessment questions could be the solution. Finally, clarity with respect to the empirical and numerical relational systems from a measurement theoretic viewpoint is a must before anyone can sensibly discuss imprecision of inputs or outputs.

A-3.7.2 Methods for Utility Assessment

The theory of methods for utility assessment has been described for MAUT in particular (Section A-3.4). The following sections describe procedures for applying several techniques including MAUT.

A-3.7.2.1 Hierarchical Utilities. This section describes two commonly used approaches to hierarchical utilities, MAUT and the Analytical Hierarchy Process (AHP). MAUT has been developed from a theoretical point of view in Section A-3.4. The discussion here will address implementation. Before discussing these approaches, however, it is important to examine the purpose of hierarchical utilities.

An interesting issue crucial to much of the discussion here involves the motivation for the use of utility hierarchies. As described, the attributes of a MAUT are those attributes that span the "space or volume" (in a vector sense) of value involved in the consequences of a decision. Usually, ten to fifteen attributes are more than enough to characterize the outcomes. More attributes mean more precision in definitions. Recall that attributes in a Decision Analytic sense are defined by observable option differences with respect to the outcomes of the option implementation denoted by acts. As the number of attributes is reduced, assuming that such reduction is for practical reasons such as number of assessments, time, etc., attributes typically will be broader possibly consisting either implicitly or explicitly of a number of sub-attributes. Thus, one reason for hierarchies is to organize dimensions of value into related clusters that can be aggregated into more general attributes. If the goal is utility assessments with respect to such aggregated attributes, the sub-attributes in the hierarchy at least provide some definition to the more aggregated attributes. However, it is important that most of the literature on human judgments points to increased difficulty and reduced reliability and validity as assessors are forced to judgmentally aggregate across several sub-attributes.

Hierarchies are also quite useful when the number of attributes involved in an analysis is large, and for purposes of summary, display, and policy analysis, aggregation is desirable. Thus, for example, numerous attributes are aggregated into a general "benefit" attribute, and many different costs are aggregated into a "cost" attribute. This type of aggregation should proceed from a "bottom up" approach where the higher level attributes become, in effect, category titles whose meaning is defined by the option differences on sub-attributes. Often several levels can intervene between high level summary attributes and the actual input assessments corresponding to actual observable or definable differences on attributes. In such cases, the relative weight given a particular attribute at any level of the hierarchy should correspond to the aggregated sum of the utility ranges of the sub-attributes that comprise that attribute as compared to a similar sum for other attributes at the same level in the hierarchy. Note that such attribute hierarchies still have "measurement purity" in that they can be precisely defined by their sub-attributes.

Still a third reason for hierarchies in utility evaluations corresponds to actual empirical usage. The hierarchy levels in some way correspond to the organizational structure or a structural decomposition of potential overall value based on general organization goals. This type of hierarchy is very common in military evaluations. Often called a "goal fabric" or "objective hierarchy", the structure is defined without the assessment of option differences. In fact, the structure can be defined, and often is, without the existence of options at all. A good reason for this is that the approach is being used to define alternatives by a top down organization of organizational needs (thus goals to meet needs). A less desirable reason for this is that often the relation of decision options to worth is not clear, and what an option would be is not defined.

Such goal fabric or objective hierarchies often are structured in terms of general desired outcomes broken out by organizational partitions such as command, geographical area, etc. The hierarchy is further developed according to organizational structure levels top down from the more general, to the specific, e.g., CINC-FLEET-TASK FORCE-OUTER BATTLE-INNER BATTLE-etc. The actual composition of the hierarchy will thus serve the purposes of representing goals, required capabilities, organizational structure, functions, etc., in an organized way that provides for a connected path from general high level goals to specific bottom level system characteristics or budget program elements. Examples of such a hierarchy approach is the Hierarchy of Objectives (See NOSC Technical Document 1938, Vol. 1) and the JDSS (see section A-5.0 of this appendix). The linkages between the observable option differences and aggregated attributes discussed with respect to MAUT, is not so clear with such hierarchies. In fact, the precision with which the relation of each hierarchy level is related to the levels above and below it varies from very poor to good enough to be operationalized in measurement procedures. Often such hierarchy developments proceed top down through a process of decomposition that attempts to capture all organizational issues, goals and objectives, and sources of uncertainty. Typically, for purposes of efficiency and clarity, such structures partition the continuum of interest at each level. A potentially very bad result is that the hierarchy becomes a huge hierarchical structure that may be in some way relatable to a decision that is to be made, but more often is not.

This issue will be labeled as the issue of "identifiability" in hierarchies, related to the general theoretical problem of the identifiability of all the parameters of a theory in an experimental test involving data collection. Is the hierarchy constructed so that the issues of option definition, uncertainty characterization and evaluation, outcome projection, and value assessment can be identified, understood, and organized with respect to decisions? If the answer is "no", then the hierarchy, though having "face validity" and organizational appeal may be of little use and actually potential damaging in policy or procurement decisions. This issue will be revisited again in discussion of methodologies, for this outcome can occur when the trade between theoretical purity and practicability swings too far in the direction of the latter.

A-3.7.2.1.1 Multi-Attribute Utility Theory (MAUT). Multi-Attribute Utility Theory (MAUT) provides analytically correct methods for assessing values with multiple effects. The foundations of MAUT were first explored by Raiffa (1969), and the first general publication of a comprehensive text on MAUT was Keeney and Raiffa (1976). In the two decades of its development, MAUT's foundations have been examined, tested, and verified by mathematicians, psychologists, and management scientists.

Many decisions, evaluations, and assessments have multiple purposes, objectives, and impacts. A decision maker is interested in striking the "best" balance among these objectives, but the concept of "best" is complex and elusive, often requiring tradeoffs among many objectives. Multiattribute utility theory provides a mathematically appropriate procedure for assessing values with multiple effects. MAUT models explicitly reflect the relative importance of each objective outcome and, therefore, the tradeoffs among them. By doing so, a MAUT model enables the decision maker to develop a summary measure or measures of value reflecting many kinds and degrees of impacts on these objectives.

The use of MAUT analysis begins with the specifications of the alternatives (i.e., decision options) that are to be evaluated. Then, the "attributes," or objectives, that are to be used to evaluate alternatives are delineated. Each alternative is then "scored" on each attribute. This scoring can be accomplished by different means. Attributes that have natural underlying quantitative measures can be scored using a single-attribute utility function. Such functions relate the underlying measure to value. Often a simple linear value function will suffice as an excellent approximation to the "theoretically correct ideal". More complex procedures may be required, however, to establish single-attribute utility functions for other attributes. (See Brown, Kahr, and Peterson, 1974, for example, for procedures for assessing utility functions that incorporate attitude toward risk. See Keeney and Raiffa, 1976, for a discussion of more complex forms of utility functions. See Edwards, 1977, for a discussion of the adequacy of linear approximations.) Scoring on attributes that do not have underlying quantitative measures is done by direct judgment. This entails establishing a scale with well-defined upper and lower endpoints and asking the experts to rate the alternatives on this scale.

The scoring of alternatives on an attribute-by-attribute basis described in the preceding discussion is formally appropriate only if some assumptions with respect to the independence of attributes are satisfied. (Keeney and Raiffa, 1976, provide the technical details of these assumptions, also see Fishburn, 1989.) Such a decomposed scoring procedure is satisfactory if the scores assigned against one attribute do not depend on the level of the other attributes. For example, if two alternatives contribute the same amount to satisfying one capability objective, then their scores on this attribute should be the same regardless of their scores on other capability objectives. This is called additive independence of the utility function. Other forms of utility functions between the totally separable additive form and the fully jointly dependent form are possible.

The next step in the development of an additive MAUT model is the specification of weights for attributes. The interpretation of weights and the procedure for assessing them depends on the form of the model to be employed to aggregate the single-attribute scores. The theoretical basis for a variety of aggregation models has been developed (see Keeney and Raiffa, 1976), but often an additive aggregation rule is appropriate or an adequate approximation, so that is our focus here. (Edwards, 1977, argues that "theory, simulation computation, and experience all suggest that weighted linear averages [the method described here] yield extremely close approximations to very much more complicated nonlinear and interactive 'true' utility functions, while remaining far easier to elicit and understand.") However, others believe that the really interesting problems are not additive.

Weights are assessed by considering the relative utility of moving from the worst to the best level on each attribute. The assessments are ratio judgments, meaning that if the utility of going from worst to best on one attribute is twice that of another attribute, the first attribute is assigned a weight twice that of the second. As earlier discussed, this assignment of weights on the basis of "swings" from worst to best is an important feature reflecting the theoretical basis of MAUT that differentiates it from weighting procedures based on some ill-defined concept of "importance." These ratio judgments are then normalized to sum to one. Normalization (or other scaling) is also a very important step that maintains the integrity and consistency of the analysis. In hierarchical multiattribute utility structures, weights are assigned and normalized at each level in the hierarchy.

The value of an alternative is calculated by multiplying the utility of an outcome by its probability of occurrence for that alternative, and summing over all outcomes. For an additive model, this can be done for each attribute separately and the result summed for the set of objective attributes. This provides the output of the process, which is the "expected utility" for that alternative. Alternatives are compared on the basis of this output.

The MAUT method has the distinction of being based on a specific objective function, the expected utility, and that it has undergone rigorous experiment of the classic procedures for eliciting utilities and probabilities. It suffers from the complexity of the measurement process and it does not give the decision maker an evaluation of the dispersion of the results, since it only provides expected value of the utility, not the spread. However, the information is available to do so.

A-3.7.2.1.2 The Analytical Hierarchy Process. The Analytical Hierarchy Process (AHP) developed by Saaty (See Saaty, T.L., 1980) has received a great deal of attention over the past ten years and has achieved a certain popularity as an alternative to MAUT. AHP is basically an approach to the assessment of utilities using a hierarchical utility process.

The AHP develops a hierarchy of goals and system functions from top level general goals and criteria down to bottom level specific criteria and even options. In discussing systems, Saaty characterizes a hierarchy as an abstraction of the structure of a system to study the functional interactions of its components. Given a general feeling that the human mind aggregates things into groups based on common attributes, Saaty indicates that the hierarchy is thus compatible with the human evaluation process. The AHP hierarchically structures the functions that the components of a system or organization are meant to serve or perform. Thus the AHP hierarchy can be a functional decomposition where the criteria at each level represent the degree to which functions are served. The AHP as used generally is well described by the type of hierarchy discussed in section A-3.7.2.1. It can have any number of levels, and the levels can contain almost any kind of objective, capability, etc. In an example in his book (Saaty, T.L., 1980, pg. 13), Saaty discusses an application involving evaluating scenarios in terms of which is best for the continued existence of the college. The levels from top to bottom of the hierarchy are respectively, focus (overall goal), forces (e.g., instruction, social life, spirit, . . .), actors (organizational components such as academic administration, faculty, students,. . .), objectives, and scenarios. The scenarios are to be evaluated. The judgment with respect to scenarios and objectives is the degree to which the scenarios determine the likelihood of achieving the objectives. Other relations include influence of objectives on actors, guidance of actors with respect to forces, and impact of forces on the welfare of the college. As we can see, relations between levels can and do vary in practice.

The AHP method develops pair-wise comparisons in which priorities are assessed for the criteria at one level with respect to impact on each of the criteria at the next level. A matrix of priorities is developed that represents the pair-wise priorities. A cell entry of the matrix is the "strength of influence" of the row criterion (lower level) with respect to the column criterion (next higher level). Such matrices should be inverse symmetric, e.g., for matrix $A = [a_{ij}]$, $a_{ij} = 1/a_{ji}$. If not, the judgments are not consistent. The procedure utilizes an exhaustive set of pair-wise comparisons, and thus can be quite time consuming, for which it has been occasionally criticized. Since any measurement procedure validly executed, including MAUT, is time consuming, such a criticism is not relevant.

Saaty has axiomatized AHP (See Saaty, 1989 and Saaty, 1977) and the representation theorem yields the basis for ratio scale utilities of the alternatives being compared. However, the relation of the axioms to the assessment procedures is not very clear due to the abstract nature of the axiomitization. The ratio scale has also been criticized for reasons which are related to the analogy to length. What is the influence of a scenario on an overall goal? Again, though this may be a problem, it is fairly insignificant compared to issues of usefulness and efficiency.

As indicated, the judgments of influence are typically not consistent. That is, the matrices are not inverse symmetric. For example, it should be the case that $a_{ik} = a_{ij}a_{jk}$ for all i, j, k. Saaty discusses consistency indicating that it is not merely the issue of transitivity of preferences, but that the actual intensity with which a preference is expressed transmits through the sequence of objects in the comparison. This implies cardinal consistency of strength of preference and generally provides the requirement and basis for a ratio scale.

The AHP method employs linear algebra techniques to resolve the inconsistencies in the matrix, and the strengths of preference for the objects in the matrix are found by determining the eigenvector of the largest eigenvalue associated with the objects in the matrix. Although not discussed, a measure of departure from consistency could be developed.

The method is thus theoretically sound from a measurement theoretic point of view. However, it has received a great deal of criticism with respect to certain characteristics. For a summary of a rather extended debate among proponents and opponents, see Winkler, 1990; Dyer, 1990; Harker and Vargas, 1990; and Saaty, 1990. Much of the criticism of the approach has to do with the problem of reversals in order that can be achieved by introducing new alternatives in the comparison set, thus violating the requirements for choices to be independent from irrelevant alternatives (those not actually in the choice set). The fact that merely adding an option in the choice set can actually cause changes in the derived ordering of options is indeed problematic from a measurement theoretic viewpoint, and the criticisms of AHP have dwelled upon this issue. Certainly if the goal of the technique were to choose one option from a set of n objects, such a problem could be very important. Also, if the set of options contains many replicas, the problem could be important. Thus in trading off issues of practicality against measurement rigor, it should be noted that the AHP does not obey the criterion of independence from irrelevant alternatives and can have reversals in order if the option set is modified. This is problematic but not yet damning of the method. A second issue involves the use of a rating scale of 1 - 9. There can be some problems with this, and it too has received a great deal of attention. The problems actually can be solved, and the argument that people cannot distinguish many more than nine categories is indeed relevant. The real issue is the linearity of the 9 point scale. The concept of a normal distribution or logarithmic scale might prove useful here for occasional, very large ratios.

This concern with order reversal and rating scale numbers perhaps obscures what is an even more important issue with the AHP, that being the precision of the comparisons. Like MAUT, the AHP, properly employed, uses ratio scale preferences and not individual numerical ratings. (Individual numerical ratings definitely would be a very serious problem in AHP.) MAUT suffered from a problem with the potential for comparing hypothetical options characterized by defined levels of attributes that were in fact descriptive of some of the options under consideration but which were not contained as part of a single option alone. Thus, preferences had to be assessed over options not in the set of alternatives for choice. In AHP, the comparisons are with respect to the greater influence of one of two criteria on the next higher level of criteria. Which object influences (impacts) the actors, say faculty, more, becoming educated or vocational preparation? Clearly these are vague judgments. Saaty in his procedural description of the AHP, discusses the importance of defining the criteria at each level, but he defines no procedures for doing so. Further, he does not indicate what a good definition would look like. Thus actual applications proceed with potentially loose or even non-existent definitions of the elements of the structure. Thus, although definition is emphasized in the description of the method, more time is spent justifying the matrix manipulations and the use of eigenvectors. We feel that the real achilles heel of this process lies in the potential for a large error in the derived scales due to imprecision of the definition of the elements of the hierarchy. This will be exacerbated if the method is used with multiple assessors, thus adding interpersonal disagreements to intra-personal unreliability (or invalidity) due to shifting criteria.

Much of the potential problem goes back to the beginnings of modern scaling and data theory in psychology when a distinction was made between extensive and intensive properties of objects. Extensive properties were those for which a physical concatenation operation could be demonstrated. Those for which no demonstration could be made were intensive. Strength was extensive, for in a test one could add weights to the amount a person must lift. Beauty was intensive. Strength of preference is also intensive.

Consider the measurement of length, which produces a ratio scale unique up to a ratio transform. Measuring length proceeds by establishing a standard sequence of measurement objects (which assumes perfect replicas are possible). Thus perfect one millimeter measures are laid end to end, thus concatenated. The object to be measured is laid next to this standard sequence and is found to be between two numbers on the scale. We then have ratio scale measurement with some error (one millimeter maximum) that can be reduced by making the standard for comparison smaller (which has been done; the standard is now the wavelength of an element).

The AHP process has no extensive properties. In general, the criteria are loosely defined if at all. An obvious improvement to most implementations would be to develop definitions of the hierarchy elements, as recommended by Saaty, on an absolute scale, if necessary, by at least giving a detailed definition and some anchor points along the postulated continuum. The bottom level alternatives should be rigorously defined as are the alternatives in a MAUT analysis. There would still be the potential for disconnects between levels of the hierarchy, but much of the potential invalidity in ratings would be reduced.

Note that this is a practical, not theoretical problem. The theoretical problems of the AHP are very minor compared to the problem with the meaningfulness of comparing the influences of two criteria on the next level objective. Unless this problem with definition is solved, all the matrix manipulation in the world will not reduce the errors involved.

A-3.7.2.1.3 Hierarchical Valuated State Space. A Hierarchical Valuated State Space is a top-down tree decomposition of objectives and situations and systems and capabilities, any of which are called states. At each branch point, there are "importance" weights normalized to add to one, and at the leaf nodes, some measure of goodness is elicited from the decision maker(s). This approach is essentially a top-down version of the hierarchical utilities mentioned above in this section.

A-3.7.2.2 Imprecisely Specified Multi-Attribute Utility Theory. Imprecisely Specified Multi-Attribute Utility Theory (ISMAUT) is a variation of MAUT that allows parameters of the decision problem to be specified "imprecisely." This imprecise specification might be an interval, an ordinal ranking, or a partial cardinal ranking. Parameters that might be specified imprecisely include scores, weights, and probabilities. (Presumably, ISMAUT could handle more general specifications of utility functions, but this could complicate ISMAUT's specification requirements and algorithms, and these extensions are not discussed in available documents on ISMAUT.)

The output of an ISMAUT analysis differs somewhat from that of a MAUT analysis. ISMAUT's primary focus is on the feasible region of the decision space. ISMAUT's output describes the conclusions that can be reached based on only the information given. While the information required for a MAUT analysis is sufficient for a complete ranking of decision options, ISMAUT requires less and thus may conclude less. MAUT requires a consistent set of weights, scores, and probabilities. This consistency is not required by ISMAUT so one possible conclusion is that there is no feasible solution set, for example, if there is no set of weights that satisfies all specifications. (In this case, procedures very much like those used in goal programming are used to identify modifications that would produce a feasible region.) In other cases, ISMAUT conclusions could include a partial or even complete ranking of options based on only the imprecise information.

A motivation for ISMAUT is the observation that some inputs required by MAUT are difficult to obtain, especially if the decision maker has not thought hard about the problem. The solution offered by ISMAUT is to allow the decision maker to provide less information. The tradeoff is that ISMAUT's procedures are much more complicated and calculationally intensive than MAUT's. ISMAUT's conclusions are also less precise. Strict ordering of the alternatives is not guaranteed. Descriptions of ISMAUT are given in White, Sage and Dozono (1984), in White (1986), and in Stewart et al. (circa 1990).

A-3.7.2.3 Fuzzy Decision Analysis Measures. The probabilities and utilities used by Watson,et al, (1979) and Freeling (1980(B)) are fuzzy versions of the ISMAUT type of probabilities and utilities. Whereas a probability or utility for MAUT is a single value, ISMAUT allows it to be a crisp interval or a crisp cardinal ranking. Fuzzy Decision Analysis allows the probability or utility to be a fuzzy interval, but does not address fuzzy cardinal ranking, although that approach could be conceived of as a fuzzy statement, such as, "p_1 is somewhat greater than p_2." Thus Fuzzy DA is a logical extension from the uncertainty of ISMAUT. But the method of aggregating results cannot be found with a linear program and the probabilistic expected utility. Instead, the fuzzy operators of max and min are used. Again, strict ordering cannot be assured.

A-3.7.2.4 Subjective Transfer Function Methodology. The Subjective Transfer Function (STF) approach described by Veit, Callero, and Rose (1984) is claimed to be "a general subjective measurement method developed to evaluate complex systems". The major underlying premise of STF, which is unstated, is that a holistic judgment is the best basis for constructing a subjective measurement scale. STF begins by developing a structure of "factors thought to directly affect outcomes of the system". There is a general confusion in STF as to whether this is an exercise to express a physical relationship, a probabilistic prediction, or an expression of value or utility. In one example, the authors show a structure for "immediate targeting" (% of important force application opportunities that could be exploited) and "target identification" (% of important enemy targets that could be identified), which seems to express a physical relationship. Another example shows a structure for "joining the Air Force" (likelihood of joining the Air Force) which is clearly a probabilistic prediction. Although not illustrated in an example, the application to a value structure is implied by the authors' contrasting STF with multiattribute utility theory and the analytical hierarchy procedure. (They also talk about structures representing "causal links" among the factors without a definition of causality.)

Next, scales are defined for each factor and discrete spanning levels are specified. Then, combinations of levels of the variables are assembled into "experimental units" and posed as "questionnaire items" to the experts. Since the responses of experts to these hypothetical combinations of factors are to be used to infer the entire measurement scale (factor scales and combination rules), a large number of questions are needed. A full factorial design is not enough. For example, an experimental design for three factors with four levels each requires 124 questionnaire items. This is a full factorial three-way design (64 questions) plus three two-way designs (48 entries) plus three one-way designs (12 questions). Many more items are needed for more factors. The expert is expected to express a numerical opinion on each question, and these answers are assumed to be "true" enough to deduce an entire value structure consisting of individual measurement scale for the three factors and the selection of the appropriate combination rule from among six hypothesized possible ones (additive, averaging, relative weight averaging, multiplicative, range, or differential weight, all of which are defined very precisely by the authors).

The authors claim to test for the right function, but the "test" is simply a visual examination of plotted data to see if the data fit together according to one of the hypothesized models (the only statistical test mentioned, an F test, is too weak to allow the desired result). Beyond this, no justification is given as to why the six combination rules are hypothesized. (Is it because these are the ones for which "tests" are available?)

The biggest weakness of STF is its basic assumption, that a measurement scale is best constructed from holistic judgments. The case for the "accuracy" of holistic judgment is weakened by the need for so many elicitations. Further doubts are raised by the manner in which some questions must be posed. For questions with incomplete information (e.g., a two-way question relating to a three-factor model), "respondents are sometimes instructed to assume a baseline capability level for factors not presented." It is assuming too much to assume that this information can help differentiate among the appropriate "interactive" functions. Holistic judgment is sometimes useful for assessing parameters within a model whose basic structure is confirmed by other means (e.g., an additive value structure) or for assessing parameters as a simplified approximation to a possibly more complex structure. Veit, Callero, and Rose (1984) are not persuasive in their arguments that STF offers a superior method for structuring and eliciting either value functions or probability relationships.

A-4.0 APPLICATION TO MILITARY WORTH

This section of the appendix will apply principles and discuss implications thus far derived to the problem of choosing military systems, which can include consideration of parts of all of the research, development, test, and acquisition decisions.

A-4.1 MODELING MILITARY WORTH - THE GENERAL APPROACH

The primary issue to be discussed is how to asses military worth. This has been characterized as a difficult, if not impossible, task, but nonetheless, one that must be accomplished. The rather obvious point is that decisions with respect to C^3 procurements are being made and indeed have been made for years. The question is whether there is a way to perform the assessment in a more effective way, with more insight into the critical elements of C3 and how it contributes to the military objectives of the overall acquisition process. The modeling approach attempts to make explicit the issues of value and uncertainty that are somehow being considered in these decisions. The decisions are made and issues of worth and associated tradeoffs are accomplished either explicitly or, more often, implicitly.) Can the decisions be assisted through the use of methods for formal decision making? If so, what is it that can be done and what are the limits on accuracy, generalization, and other relevant implementation issues? This section will address some of the answers to these questions.

A-4.1.1 The Decision Analytic Paradigm as a Standard

A major premise of this report is that the Decision Analytic paradigm, based on maximizing expected utility, is the appropriate one to use as a basis for addressing the military worth problem. It is appropriate because it is the best developed discipline for addressing complex decisions involving uncertainty and value tradeoffs and because it is supported by a well understood, tested, axiomatic foundation. This view of the appropriateness of Decision Analysis in the narrow sense is not meant to exclude any other techniques or approaches. Nor is it to exclude issues and tradeoffs relevant to decision analysis in "the wider sense". These will be considered as appropriate. Rather, the Decision Analytic paradigm will be used as a basis for addressing issues of uncertainty and value and for comparing other approaches.

As discussed earlier in the appendix, a decision is viewed as an act followed by subsequent events and other acts that unfold over time eventually terminating in a set of outcomes that can be characterized as outcome levels on attributes for which the decision making organization has value or utility. The problem then for a particular decision is to characterize the acts, events, and attributes, to structure them appropriately, to determine or elicit valid, possibly subjective, conditional probabilities and utilities, to characterize the capabilities of the alternatives, and to assess the alternatives on the basis of how their capabilities affect expected utility of the decision objectives.

A-4.1.2 Developing the Evaluation Structure - Some decisions

As earlier discussed in the appendix, military evaluations often utilize hierarchies to characterize worth. The Decision Analytic paradigm employs MAUT for assessing the worth of a complex outcome, and the decision tree consists of a large set of branches with terminal "leaves". The decision tree can be viewed in a sense as a means of structuring the prediction of the attribute levels conditional on choice of act. The problem of relating the act to the eventual probable outcome is decomposed into appropriate, successive, conditional judgments culminating in preference tradeoffs over attribute levels. This is the appropriate prescriptive framework for assessing military worth, for the prediction of future worth is appropriately structured into an axiomatically based set of conditional assessments with respect to uncertainty and value.

How then does the use of MAUT hierarchies relate to the Decision Analytic paradigm as a standard for worth assessment? Many attempts to measure military worth have employed MAUT hierarchies in various forms. (See for examples Hayes, M.L. et. al., 1975; Zeller, G.A., 1989). More generally, military evaluations have employed worth hierarchies which have higher levels consisting of aggregate attributes of worth such as overall mission achievement. The hierarchies are employed to successively decompose these aggregate attributes into the more specific levels over which they were aggregated. The bottom of the hierarchy usually has fairly specific attributes such as capability objectives (e.g., in JDSS), an actual system performance characteristic, or a specific rating with respect to a criterion, e.g., technical suitability with respect to sea control in a Mediterranean scenario. The difference between such a hierarchy and the decision tree is important. The decision tree involves folding back of utility over attribute levels, and probabilities on those levels, to an act thus yielding the subjective expected utility of the decision to act. This is different from rolling up a hierarchy from specific attributes to general attributes. Before discussing theoretical and practical meanings of the difference, more attention will be given to construction of such hierarchies.

A-4.1.2.1 Constructing Worth Hierarchies. Hierarchies can be constructed top down, bottom up, or both. In practicing the bottom up approach, a set of attributes is created, usually by examining differences in alternatives. Then the attributes are grouped by similarity or by some relation in a prediction criterion (prediction of overall worth). Then at a second level, attributes are formed as combinations of the bottom level attributes and so on up through the hierarchy. As this is done, the second level attribute formed represents a volume in a value space. That is, the attribute characterizes that volume obtained by swinging the bottom level attributes that comprise it from the point corresponding to their joint minimum to the point corresponding to their joint maximum. It should be noted that the terminal attributes of the decision tree could be grouped hierarchically if need be, but every attempt is made to keep the number of attributes small. The attributes are major attributes of worth, and the decision tree has provided a very specific set of details (i.e., a branch) conditional upon which the required assessments can be made.

Top down construction of hierarchies begins with some overall measure of worth such as "overall force effectiveness" or "total organizational worth". A hierarchy of attributes is formed where the attributes of one level are successively decomposed into sub-attributes at the next level until a bottom level is reached at which it is deemed feasible to assess direct option rating or preferences with respect to the bottom level attributes (over the options being evaluated). Usually the numbers of levels are not symmetric across attributes. Also, in some cases, there are no specific options to be evaluated. Rather the hierarchy is a hierarchy of requirements or gaps, e.g., functional requirements or potential functional deficiencies. Options are then the combinations of various levels of requirements satisfaction achieved by the alternatives. Trade-offs are offsetting differences in the utilities of requirements variables. One option may be higher in utility of one variable and lower in the other. When the utilities are additive, these trade-offs are comparable for each variable.

But neither of these hierarchies of weighted worth are relevant to the question. Only the high level attributes are of utility to the decision maker. Lower level attributes do not contribute in a weighted sum way to the higher level ones. The relationship is a model of the effects of the lower on the higher. These hierarchies are not actually part of the MAUT theory; rather, they are heuristic work-arounds for a gap in the application of the theory. That gap can be filled by conditional modeling using causality nets or influence diagrams.

A-4.1.2.2 Assessing Values Within the Hierarchy. Hierarchies of the type discussed usually involve assumption of additivity, often with minimal checks. (If obvious violations of additivity occur, they can be included as side conditions on the calculations in the analysis or on the conclusions to be drawn.) Assuming additivity, intra-attribute value functions are assessed for each bottom level attribute of the hierarchy (often termed the "input score" level in applications). Then intra-attribute weights are assessed by comparing the values of swinging attributes through their ranges. Bottom level attribute weights are assessed by comparing single attribute value swings, but unlike these comparisons, the higher level attribute weights correspond to the simultaneous swing of multiple attributes through their respective ranges. Such worth differences for aggregated attributes become more difficult as standards for comparison.

One practice used in building such hierarchies is to provide the attribute with a name associated with the concept with respect to which the grouped attributes are similar, e.g., operational suitability in desert terrain, logistical supportability, vulnerability, etc. Weights are then directly assessed with respect to the range of this concept as evidenced by the alternatives in the analysis. The problem is that the standard for assessment can be unclear here unless carefully defined and explained to the experts providing the assessments. A good example of this problem, often encountered in military evaluation hierarchies, is the evaluation attribute, "system reliability". It is often decomposed into or aggregated from "mean time between failure" (in hours) and "mean time to repair" (sometimes decomposed into "in the field" and "in depot" (including time to get it there)). The experts are asked to assess the worth of increased reliability versus increases in the range of the platform or some similar capability. Obviously, there is a potential for inconsistencies both within and across experts due to imprecision in attribute definition. Also, the aggregate factor may be associated with value differences that are related to the general category associated with the name and not with the actual attributes in the analysis nor the cause and effect of either reliablity or range on the military outcome.

A second method for assessing weights given appropriate intra-attribute value functions (and satisfaction of additivity) is to assign each aggregate attribute the sum of the weights associated with its sub-attributes, and thus corresponding to actual potential value swings in the analysis. This approach to weighting is also often called "bottom up weighting" to distinguish it from the previous approach usually employed in "top down" analyses. It is normatively correct given additivity, and is quite useful for displaying aggregate value comparisons (e.g., benefit attributes versus cost attribute kinds of comparisons). Often the attributes have resultant weights that do not correspond to intuitive feelings about such attributes due to a difference between the general population of alternatives and the specific alternatives in the analysis. This is a significant problem because the results should not be dependent on the collection of alternatives. The problem is due to the lack of foundation for this kind of hierarchy.

A-4.1.3 A Major Issue - Prediction of Performance and Assessment of Worth

Section A-4.1.2 has described developing a worth hierarchy and assessing weights as often practiced. It has been noted that this approach may not be the same as employing the Decision Analytic paradigm. The issue has to do with uncertainty versus value. In the decision tree, the entire structure is defined to inter-relate the uncertain variables in a manner that provides for valid conditional assessments of worth (or utility). The worth calculated for an act in the tree as the tree is rolled back is a conditional expected utility. The tree is structured based on issues of utility additivity at the leaf nodes, and the major part of the structure has to do with inter-relating the acts and events. This tree is not the tree of worth hierarchies; it is a tree of probabilities. The utilities only appear at the leaves. The tree is a chain of conditional probabilities, but this is usually not recognized, and the conditional independence is not exploited.

The choice of a system involves estimating its future worth. From a measurement theoretic point of view, there is not a set of objects with some preference and thus an observable relation to be assessed over the objects (e.g., systems or combinations of capabilities, or even single capabilities), for, then, any preference would not involve the actual outcomes. Rather an implicit prediction of performance and associated worth would have to be made over the capabilities to be chosen. The decision tree serves as a means to validly represent the predictive aspect of the problem using the Bayesian statistical approach as well as characterize the potential worth of the outcomes. To what degree will the capabilities we choose now yield the outcomes (attribute levels) we wish to achieve in the future?

Worth hierarchies fail to answer this question due to varying degrees of validity depending on how much the prediction issues and value assessments are intermingled. Certain hierarchies mix scenarios at some levels with lower level capabilities at others. Others use "influence" in employing the AHP or a similar process. Others actually use the same logic as decision trees and build conditional probability hierarchies (an obvious example being the Conditional Probability Logic - CPL). Consider a bottom level of a hierarchy to be a physical characteristic of a communication link in terms of throughput in bits per second. One judgment could be to express a preference for levels of this versus levels of sensor capability. A second judgment could be to estimate the potential worth of this in combat operations in a defined combat scenario in the Middle East. Still a third would be to try to assess the conditional probability that this level of capability would yield the required communications continuity to support the combat operations in a scenario to produce a specified conditional outcome. These are different judgments. One is a judgment of the worth of the outcome. Probability judgments are made with respect to estimates of performance as related to those outcomes. Both are required, but when worth assessments serve as surrogates for performance predictions, severe validity issues can arise. Worth judgments must be made at the level of commonality of purpose. A sensor and a communications system have different direct purposes. Their common purpose is the achievement of a higher level attribute, like process the sensor data at another location. The worth must be assessed at the process output attribute outcome, such as, data fusion events.

One issue here is value-wise independence versus stochastic independence. There has been a great deal of literature (e.g., Edwards, W., 1977) suggesting that violations of additivity in utility assessment are not serious, and this argument is usually based on the adequacy of a linear multi-valued function as an approximation to a non-linear combination of the same values. For cases of simple value, this is often true. However, the problem here is more complicated, for we are not just trying to approximate a non-additive worth, but rather we are trying to understand and predict the relationship of a combination of capabilities to the eventual effectiveness of a military force, and to assess a value for that effectiveness. The model relating the low level capabilities to the higher level attributes is a probability function, not a utility weighting function.

It is acceptable in a non-linear utility process to assess values over attributes, which are, in fact, dependent, as if they were independent, thus incurring some error. It is less acceptable to treat low level events as attributes and combine them using MAUT, without modeling their effects on the high level attributes. An example may clarify this.

In choosing a new position, I can construct an additive evaluation function in which I independently assess a value function over distance from home to work as one attribute and another function over type of work. It does not matter that the type of jobs I like are all associated with locations with long commutes. This simply means that the set of options which result in outcomes that I value most highly, e.g., my most preferred kind of work at a short commute from home, is, in fact, empty. However, consider a military system in which I consider the range of a weapon and the probability that it will neutralize or destroy the target given a hit. Suppose further that the targets of most concern occur at long range and are difficult to destroy, thus requiring high lethality. If additive utilities are assessed over these range and lethality variables, the function will assign moderate value to options that have potentially no use at all. Of interest is the destruction of the target. The probability of this outcome is dependent on the joint occurrence of levels of the two variables, and the joint probability of their occurrence.

We could state that the error incurred here would be small and thus acceptable as with the job choice, but there is a major difference. With the job choice, the alternatives will not be defined by the value function developed. Rather, the value function will help in choosing among alternatives that are candidates for evaluation. In the military example, the choice of capability level desired (required) could be determined by the value function of the high level objectives, when reflected in the model of the effects of the capabilities on the high level attributes. If the non-additivity here is ignored, and if enough of these types of cases occur, the resultant additive value function could assign high worths to combined capability levels that are in essence worthless. The additive utility approach could conclude that if we choose a low level of range with a maximum lethality, this combination is equivalent to threshold levels (defined as some level along the continuum) in both. A probability model would assign the desired outcome zero probability for the first combination and some moderate to low, but non-zero, probability for the second combination.

Again, the problem is combining low level events as if they were attributes over which value is to be assessed. The problem is further exacerbated if the results of the analysis have to be explained. It is one thing to state that an approximate model has been used and that the error is expected to be so much. It is quite another to decompose overall values into sub-attributes to identify "drivers" of the results which are, in fact, erroneously combined.

A-4.1.4 Comparing Bottom Up Versus Top Down Approaches

The previous section discussed problems with using a hierarchy of worth as a substitute or an approximation to a probability model of the decision tree. This section will briefly assume that a hierarchy of worth is to be utilized and will compare the bottom up and top down weighting procedures.

In the bottom up procedure, bottom level attribute comparisons are made by comparing intra-attribute range swings in terms of utility. Higher level attribute weights are then derived from the sums of the weights of sub-attributes that comprise the attributes. The process continues in a roll up to the top, normalizing weights at each level. Recall that the intra-attribute weights may not make intuitive sense in that they do not reflect intuitive judgments about overall attribute importances. There are several reasons for this.

One reason is that bottom up weights are derived from differences among the options being evaluated, and these may not form a representative sample of the aggregate attribute ranges. Thus some attribute ranges of utility will, in effect, be truncated. A second problem can occur in such hierarchies when attributes are decomposed into differential detail with respect to sub-attributes. Many attributes, especially technically oriented ones, can be characterized by parameters and are often broken into many sub-attributes so as to exhaustively characterize the attribute. Thus, if eight system aspects are involved in determining an operational outcome, all may be included. They may be highly correlated in a statistical sense, and here the problem of confusing prediction with valuation can again be an issue. There is a tendency to give some weight to any parameter in the model. Thus, the most important sub-attribute in a group is assigned an arbitrary anchor "score" of 100. Other sub-attributes are given "scores" which are in effect insufficient adjustments down from 100. The result is an over-weighting of the attribute when the sub-attribute scores are aggregated. This may be countered somewhat by a combination of top down and bottom up weighting.

Top down weighting tends to be independent of the actual range of utility involved in the decision. Thus attributes are compared with respect to some universal criteria or options, and importance or utility is partitioned at each level, successively decomposing so that at each level, the relative contribution of sub-attributes to the next highest level is reflected by the partitioned benefit assessments. At the bottom level, the criteria upon which options are actively compared are thus given weights derived from the potential relative contribution to the next highest level, as opposed to observable utility differences among the options.

Another issue arises when there are no specific options to evaluate. Suppose the hierarchy has, at the bottom, fairly specific attributes, which are, in fact, capability objectives or functional capabilities. There need not be any specific systems representing alternatives. Rather, each alternative might be a combination of levels of bottom level attributes representing some specification of requirements for a notional system. The Decision Analytic paradigm is still relevant for appropriately structuring event and value dependencies. But, if the set of attributes consists of attributes that depend on other attributes, a hierarchy of requirements might result. Such structures, alternately called goal fabrics, requirements hierarchies, objective hierarchies, and the like are fairly common and, depending on how utilized, can be subject to the problems thus far discussed in this appendix. One approach to this problem, however, combines the hierarchical decomposition approach with probabilistic combination in a manner that is designed to avoid some of these pitfalls. That approach will be described in the next section.

A-4.2 MODELING WORTH USING THE HIERARCHY OF OBJECTIVES AND CONDITIONAL PROBABILITY LOGIC

The concept of decomposing an overall objective into sub-objectives that are more and more detailed was discussed in Section A-4.1. Work by Girard (See Girard, 1989B, as well as NOSC Technical Document 1938) utilizes a hierarchy of objectives to characterize command and control warfare requirements. The measure of attainment of such objectives or satisfaction of the requirements is developed as the decision probability using conditional probability logic. Thus the distinction between prediction and valuation is made and many of the problems discussed thus far are avoided.

In the Hierarchy of Objectives terminology, mission objectives are based on achieving a preferred set of outcomes which correspond to specific states of the enemy and friendly forces and also other defined variables. For example, where deterrence of combat is preferred, the political state in a country would be directly related to the combat readiness and would thus be a potential outcome. The military forces employ strategies and plans represented by a hierarchy of functions that correspond with the Hierarchy of Objectives. For the Hierarchy of Objectives, there is a Hierarchy of Sub-objectives and a parallel associated Hierarchy of Functions and Outcomes of function execution, the latter being directly related to the sub-objectives. The purpose of the Hierarchy of Objectives (as stated in Girard, 1990) is to define the set of functions and their favorable (and unfavorable) outcomes at each level, including overall mission level, sub-objective or functional level, and system capability level. The outcomes form the basis for defining measures of potential achievement of the objectives using conditional probability logic.

The Hierarchy of Objectives assessment process relates to overall mission success and is thus affected by weapon system capability, sensor capability, other resource qualities, troop training, and many more identifiable force characteristics. C3 is important to glue all these together in implementation, and it is done through decision making. Decision making is performed in association with all objectives and thus occurs at all levels of the hierarchy. The purpose of decision making (defined by Girard, 1990, p. 3) is "to allocate resources to perform functions in support of higher objectives." Decisions thus identify objectives to pursue, functions to perform, resources to use, times, and the like. Girard distinguishes between interpreting information and choosing courses of action, both called Command Decisions. C3 functions, therefore, support the implementation of command decisions which in turn, control the behavior of the system at all levels. C3 includes the communications functions and information processing and display and associated outcomes determine the availability of information required for command decisions. These outcomes are conditional on the performance of other functions, e.g., surveillance, intelligence, etc. The command functions then enable (and direct the extent or specific parameters of) the performance of other functions at each hierarchy level. This terminology provides the link to conditional probabilities in the Hierarchy of Objectives logic.

Note that different approaches to estimating force effectiveness or evaluating system capability often use either simulations or hierarchies of attributes. Usually, the two are not mixed. Simulation results can be difficult to use in decisions because only performance measures specific to the simulation are available. Traceability to achievement of hierarchy sub-objectives is not generally a capability of the simulation. Further, as earlier indicated, integrating C3 evaluation with evaluation of other system components has been difficult. The Hierarchy of Objectives approach provides a means of integrating evaluations of sensor, weapon, C3, and other capabilities by examining the joint conditional probabilities of functional outcomes. One could envision the integration of the output of a highly detailed weapon system's performance evaluation with a partially subjective evaluation of the probability of a particular command decision outcome in an evaluation using this approach. Such a demonstrated capability has been greatly in need as indicated several times in this appendix.

The evaluation of C3 using the Hierarchy of Objectives approach uses decision probability at all levels, with respect to achievement of very low level outcomes as well as higher level outcomes related to general mission objectives. Decision probability thus becomes a common measure at all levels of the hierarchy. However, as noted by Girard (1990, p. 9), requirements in general are not stated as conditional probabilities, and to use such probabilities as a common measure may require the translation of other requirements into probabilistic terms. This can often be done in terms of confidence of achieving a certain level of performance, e.g., the probability that a sensor of Type S will be able to detect targets of Type T under conditions specific in Scenario C. C3 requirements can be stated in terms of probabilities, as are the following examples of operational C3 requirements (See Girard, 1990):

P (Data Obtained/Communications connectivity, status of resources)
P (Accurate Picture/Data obtained, data accuracy, time delay, prior expectations)
P (Recognize Situation/Accuracy of Picture, expectations)
P (Course of Action (COA), resource selected/Plan in place, situation as recognized, authority to act)
P (Decision promulgated/COA, communications, connectivity, etc.)

The scenario then unfolds in terms of other conditional functional performances. As indicated by Girard (1990, p. 10), "with these types of C3I measures in place, along with the Mission-oriented requirements measures, the overall outcome can be assessed with a view into the contribution of C3I, since the activation of Warfare Mission Area functions is conditional on the direction to carry them out." Girard further discusses how to incorporate timeliness into such probabilities, as it is not only the decision quality but also the timeliness that is relevant.

The overall approach to evaluation of C3 requirements using the Hierarchy of Objectives and Conditional Probability Logic (CPL) is, of course, more complicated than explained here and is explained in detail in several references. (NOSC Technical Document 1938; Girard, 1989(A); Girard 1989(B); Girard, 1990) The important consideration here is the use of the conditional probability logic and functional outcome interrelations to characterize C3 performance. The relation of this approach to Decision Analysis and other approaches as well as some potential implementation issues will be briefly discussed.

A-4.2.1 Methodological and Implementation Issues with the Hierarchy of Objectives

Discussion of the Hierarchy of Objectives and Conditional Probability Logic approaches in the previous section has been brief, focusing mainly on the basic theoretical approach. As indicated, the actual approach is very detailed and is explained quite clearly in the aforementioned references. The complexity of the approach is more a property of the domain of the problem and not the approach. As indicated earlier, many have claimed that a valid solution to the evaluation problem (from either a theoretical or practical viewpoint) may be impossible. The approach elucidated by Girard (1989, 1990) systematically attacks the difficulties of the problem and lays out the methodological framework.

The representation of a problem can be extremely complicated depending on the degree of fidelity to which the functional performances and interrelations are represented. The degree of complexity employed in the Conditional Probability Logic relates to an earlier distinction which has been made between decision analysis in the narrow and wide sense. This distinction is, in fact, an arbitrary one attempting to partition errors into errors due to implementation of assessment techniques and errors that result from practical tradeoffs with respect to empirical problem constraints. Thus, Decision Analytic errors in the narrow sense would include miscalibrations on probability distributions, unreliable intra-attribute utility assessments, and other measurement type issues.

An example of an error in the wider sense would include using a simplified problem framework which includes four or five scenarios to represent the potential future of operational engagements when it is clear that these scenarios represent a gross cut at representing an unknown future. The amount of error involved would remain undetermined as would the ability to indicate the impact of the approximation on the decision. The degree to which the error in such an approximation is calibrated is important and is part of the "continuum" of errors or tradeoffs between the narrow and wide sense. The point is that a deliberate decision is made to approximate a valid framework as a representation, possibly without knowing how much error is involved. How bad is this?

Statistical decision approaches typically characterize the unknowns as random variables thus allowing specification of an error distribution and statements about statistical significance and errors of various types. The error incurred by approximating a non-additive problem structure by an additive one can be specified if the non-additive structure can be specified. If it is left as a diffuse alternative to the representation chosen, then the error of approximation cannot be quantified. This is a common problem in evaluating military worth. The "correct" model to "validly" represent the "truth" is far too detailed to even attempt to employ. An example would be a decision tree to represent a decision to employ a particular combination of C^3 capabilities. The number of potential acts and events even if confined to those contributing most of the variance of outcomes would result in a prohibitively large decision tree that could not be practically utilized.

The Conditional Probability Logic approach also provides a means to represent pieces of the problem at differing levels of detail. Certain parts of a problem can be simulated to yield conditional probabilities of objective achievement while other parts can be straight subjective assessments of conditional probabilities based on pieces of evidence with respect to functional performance. This provides a means of attacking the entire problem without an unmanageable explosion in detail.

<u>When the Hierarchy of Objectives Model, implemented with Conditional Probability Logic, is combined with Multi-Attribute Utility Theory, a Model of Choice Under Uncertainty is the result.</u> The performance (and cost) models are inserted between the set of alternatives, including their capabilities, and the top level objectives, with utility functions applied to the important attributes associated with those objectives.

A-4.3 OTHER MODELING ISSUES IN THE ASSESSMENT OF MILITARY WORTH: SOME ISSUES WITH SIMULATIONS

Several other approaches to the assessment of worth will be briefly discussed for purposes of completeness. One method that has been discussed is simulation. Simulation models vary greatly in terms of complexity. Simple analytical models can be developed real-time by developing a set of equations relating relevant force parameters on both sides such as that done with Lanchester equations where force attributions through casualties are essentially a function of force size, and these are calculated using simple equations. The models can be made more complex by introducing simple factors such as maneuverability, logistics support, etc., and other factors which serve to modify the critical parameters, but the models are analytical models involving tradeoffs with respect to key model parameters. Another complication is to introduce randomness through random variables that are dependent on specified sources of uncertainty. Large simulations are built based on the same principals and the numbers and types are too many to discuss here. They have been used extensively to model weapon, sensor, system, and force performance. They can employ models that are based on engineering principles, physics, electronics, system dynamics, and numerous other disciplines, thus attempting to capture all the aspects of the problem that are relevant to determining variation in performance. In terms of doing so, they are often admirable creations. In terms of what to do with them, decision makers often have difficulty.

A-4.3.1 Simulating the Right Problem

Simulations, to be useful, are often quite complex, and the details of all the interactions can be quite difficult to grasp. As model inputs are varied and outputs observed, the decision maker is provided with certain tradeoffs. A typical tradeoff would be with respect to such parameters of an aircraft as speed, payload, lethality, range, etc. As these are varied, different questions can be answered relative to a specified combat situation.

With weapons and sensors, simulations have been fairly successful and useful. Part of the reason is that the interactions among weapons and sensors obey the physical and engineering principles that are well understood and for which sophisticated modeling technology exists. Thus, if one wishes to model whether to consider naval battle groups using conventional carriers with conventional aircraft (e.g., F14's) versus carriers designed for and equipped with Vertical Takeoff and Landing (VTOL) aircraft, the simulation approach can be employed to address many of the questions. Engagement scenarios can be developed in which these naval battle groups are engaged in different types of situations. The VTOL air systems will have shorter ranges because of the requirements for high fuel expenditures during takeoff and landing (if they employ that mode, which could be optional). Thus, range becomes a critical parameter. Perhaps the VTOL planes are more maneuverable and maneuverability becomes another parameter to vary. Perhaps a well-developed, well-understood simulation can be employed.

However, this simple example points out one problem often encountered with simulations. Because the simulations are complex, they have historically been expensive to develop and to run, a trend which should be changing with the rapid developments in computer science. Thus, they are often used to address problems for which they were not specifically designed. The problem is "force fitted" to the simulation. This can be very dangerous. Because the simulations are complex, the decision maker can be at the mercy of the analyst. Depending on how questions are asked, miscommunications can occur.

Returning to the VTOL versus conventional force example, a simulation could be conditional involving several different types of engagements that vary in terms of requirements for range, maneuverability, system lethality, etc., and conclusions drawn. But how useful would such conclusions be? Consider some questions. Why use a VTOL system? What are the advantages of a VTOL system? Answers such as catapult removal on carriers, elimination of fouled carrier decks, elimination of requirements for direction changes with respect to the wind, and the like become relevant. How can these advantages be compared with disadvantages such as those due to reduced range? An experienced analyst would reject subjective assessments by experts indicating that the problem is too complex. A critic of simulations would indicate that the available simulations do not address these system differences critical to actual differences in force effectiveness. What is the decision maker to do?

The first thing he must do is understand what the simulation can and cannot do. If the simulation as it stands cannot address the differences, perhaps fairly simple modifications can be made to incorporate the systems differences. In the VTOL example, many of the VTOL advantages affect sortie rates, i.e., air system turn-around and carrier vulnerability. Perhaps modifications of the simulation can account for most of these.

Another approach is to use the simulation results caveated by the degree to which the actual situation is approximated. The simulation model output can be used as one piece of evidence in an analysis based, say, on the Conditional Probability Logic. Other pieces of evidence possibly including subjective assessments could be combined with the simulation outputs to yield realistic assessments of potential performance.

A-4.3.2 Simulations and C^3: Difficulties in Tradeoffs

The discussion in the previous section related to simulations in general, indicating a potential problem with addressing tradeoffs that are not prespecified and also, a lesser problem, in understanding the causal details in performance tradeoffs. Another problem is the issue of C^3 in operational simulations. To what degree can the effects of command decisions, control capabilities, and communications capability be represented? As indicated in the Conditional Probability Logic discussion, the probability and timeliness of a command decision (as well as its quality, i.e., probability of choosing the best or optimal option) need to be assessed. Similarly, the probability degree and quality of system and force control are to be addressed. Similarly, communications throughput, in terms of message receipt probability, and timeliness in terms of probability of receipt in time to implement an appropriate response, are also important. The degree to which C^3 issues can be validly represented, however, is a serious issue in most applications. How can the impacts of C^3 variables with respect to weapon, sensor, and overall force effectiveness be assessed? How are the effects of organizational structure assessed? How are the impacts of humans in the loop characterized? These are not new questions. What is offered here are some concepts with respect to systematically addressing such issues that provide some validity from a decision analytic viewpoint, both in the narrow and wide sense.

Consider a simple question faced in a force structure evaluation. Should the command structure be centralized or de-centralized? What are the effects with respect to command decisions, control capability, and communications? Clearly such issues in situation assessment (and all inferential assessments required) could be impacted. Connectivity of communications, survivability of key nodes, graceful degradation, and other issues become key. Note that performance prior to any force damage includes interesting tradeoffs and can be addressed. However, many of the major questions here, and reasons for choosing one structure over the other, concern survivability, graceful degradation of performance, and other issues involving partial force effectiveness. The ability to validly, simultaneously address such tradeoffs has not been demonstrated. Which would be riskier -- to have a centralized command node that could be lost, or a more survivable, de-centralized command structure where the complete picture would be difficult to assess and total force planning could be of lower quality? Could the overhead of a hybrid structure be too great or is this a sensible alternative -- say a centralized structure that degrades to de-centralized with pre-defined conditional response to "scenario templates"? Such questions with respect to C^3 are important and are extremely difficult to model. Further, subjective assessments with respect to them fast become too difficult for the human processing capability. But decisions concerning which C3 systems should be acquired depend on these questions.

Again, the Hierarchy of Objectives and the Conditional Probability Logic provide a means of decomposing the problem into potentially manageable sub-parts. In such cases, the effect of approximations could be pinpointed, and sub-parts could be modeled or could be a basis for conditional assessments. The same would be true using an integrated network of semi-independent influence diagrams connected by key event nodes.

A-4.3.3 Petri Nets, IDEF, and Other Functional Models

Certain simulation tools do exist for modeling C^3 networks and have been employed successfully. For good examples see Small and Groveston (1985), Levis and Boettcher (1983), and Levis and Monguillet (1988). The Small-Groveston work uses the IDEF structured modeling language. The Levis et. al. work uses colored Petri Nets as a way of modeling command and control networks. Petri nets are a special case of the more general causality net described by Girard (1990) and NOSC TD 1938, vol. 2.

The Petri Net approach to the analysis and design of information-processing and decision-making organizations has been developed by Levis and his colleagues based on the model of interacting human decision makers (DMs) with limited information-processing capabilities, characterized by March (1978) as bounded rationality. This theoretical approach makes use of Petri Net diagrams for modeling command and control structures. Petri Nets have been used extensively in representing and analyzing computing systems and processes. Using a small set of system primitives and the corresponding Petri Net elements, a decision-making model can be converted into an equivalent Petri Net. Decision-making organizations (DMOs) can be represented by a Petri Net subnet, as shown in Figure A-3, with four functions or transitions (shown by vertical bars) and three internal places (shown by circles). The model allows differentiation among inputs and outputs of the decision maker, and also allows description of the types of interactions that can exist between two decision makers.

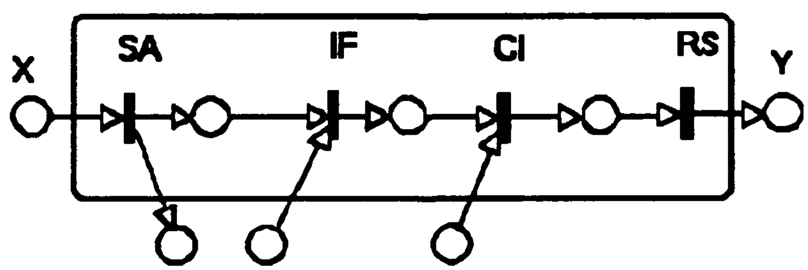


Figure A-3. Petri Net of the Interacting Decision Maker

The SA function in Figure A-3 is the situation assessment function. The DM receives an input signal X from an external source (one signal at a time assumed). The DM processes the signal with a specific algorithm which matches X (with some probability) to a situation the decision maker already knows. The result is an assessed situation which he can share with other decision makers. The next function is information fusion (IF). He sends the SA result to the IF function and also can receive inputs from other DMs' SA functions at this point (information sharing). He can also receive the results of another DM's entire decision-making process at the IF stage (results sharing). He fuses all the inputs to the IF function to yield the final situation assessment. This is sent on to the command interpretation (CI) function in which consideration is made of commands from other DMs which could result in a restriction of his set of alternatives for generating a response to the input. An external input such as that shown in Figure A-3 to the CI function can only be from the final output (response selection) of another DM. This implies a subordination of this DM to the one issuing the command. The output of the CI stage is a command used in the next stage, the response selection (RS) function which produces the output Y of the decision maker. Note that Y can be sent as input to another DM's SA or IF functions (results sharing) or his CI function as a command. These are the possible interactions between DMs using this representation. Of course, other functions and related interactions could be added.

In this representation the transitions (vertical bars) correspond to algorithms, the connectors to precedence relations between the algorithms, and the tokens (dots in circles) are the inputs and outputs of the algorithms. An algorithm can run only when all its token places (circles) are occupied and the time it takes to run is called the transition time. In both the hierarchical and parallel organizations to be discussed, the SA and RS stages contain several algorithms and a decision switch determines the choice of algorithm. The switch position is determined by the decision strategies of the decision maker. Thus, the strategies are the decision rules the decision maker uses to adopt a specific algorithm for a specific transition (function), e.g., situation assessment. These strategies can be deterministic or stochastic, and can be based on the attributes (colors) of the input token or be independent of these attributes (non-colored).

The evaluation of the Petri Net approach to modeling organizations has compared parallel and hierarchical organizations (Andreadakis and Levis, 1987). Figure A-4 presents Petri Net representations of a parallel and a hierarchical decision-making organization that will be used in discussing an air defense problem to be presented. The separate functions in these organizations are not labeled.

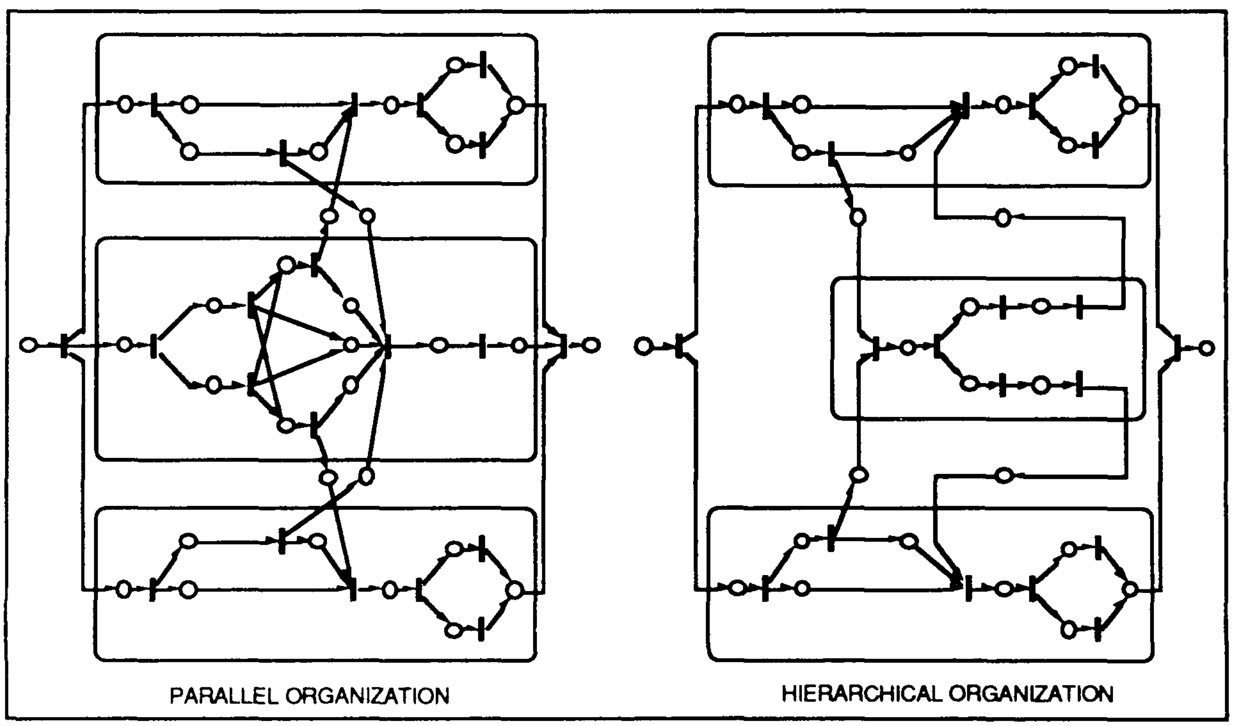


Figure A-4. Petri Net Representation of Parallel and Hierarchical Organizations

In each organization there are three decision makers represented within the enclosed rectangles. The incoming information is represented by the circle for the signal to the left of the diagrams. In the parallel organization, the signal is sent to all three decision makers. In the hierarchical organization, it is sent to the decision makers represented by the top and bottom rectangles. The vertical bars in the organizations represent functions, as previously described. A signal is processed by the function and the transformed information is sent on to one or more other functions within the decision-making organization, or possibly to a function in another decision-making organization.

If we assume an air defense situation, in the parallel organization the airspace to be defended has been divided into three sectors with each decision maker assigned to one sector. Each decision maker can observe and engage threats in his sector. However, threats can move between sectors and thus information must be shared between decision makers in adjacent sectors. Thus, the middle decision maker must communicate with both neighbors and has the highest workload. Each decision maker at some time sends the results of his situation assessment to the information fusion function of the adjacent decision maker.

In the hierarchical organization, the airspace is divided into two sectors and each sector is assigned one decision maker. A center region straddles the two sectors and a supervisor is introduced. The Petri Net representation in Figure A-4 is a model involving the decision makers, not the physical sectors. The middle decision maker in the hierarchical organization commands the central region which straddles the outer two sectors. The supervisor does not observe the airspace directly, but he does receive information about threats in the central region from the other two decision makers. He processes information provided and then allocates the threat to one of the other two decision makers depending on the information provided. Because he issues orders to the other two decision makers, he is in a superordinate role. This is one form of hierarchical organization. In this case, the center decision maker sends the results of his decision as a command to the command interpretation function of one or the other adjacent decision makers.

Several measures of performance have been used to describe the performance of organizations such as these. These measures are key to the evaluation of team decision-making performance.

_*Accuracy of Response (J)*: How well do the outputs of the organization match the desired ones?

_*Timeliness of Response (T)*: Are the responses within the appropriate window of opportunity?

_*Individual Workload (G)*: Can each decision maker perform his share of the task without exceeding his bounded rationality?

_*Consistency of Information (D)*: The extent to which information can be fused without contradictions.

_*Synchronization (S)*: The amount of time decision makers have to wait to receive the needed information in order to proceed.

_*Coordination*: The execution of a process is coordinated if it is consistent and synchronized, that is, if information can be fused without contradictions, and information is available to every decision maker when he is ready for it. A task execution is coordinated if its component processes are coordinated. An organization is coordinated if all its tasks are coordinated.

These measures are further defined in terms of Petri Net symbology. For example, the firing of a transition (function) is synchronized if and only if all the enabling tokens (arriving information) have the same task arrival times and their local arrival times (at the local transition) are less than the deadlines for that transition. The firing of a transition is consistent if and only if all the enabling tokens (arriving pieces of information) have the same task arrival times and belong to the same class C.

The above measures are useful because they are clearly desirable properties of the response of any decision-making organization. At the same time they can be assessed using the model-driven experimentation approach. They also lend themselves nicely to studies of team decision-making performance. Note that consistency, synchronization and coordination are more difficult to measure than accuracy and timeliness. However, accuracy and timeliness are required for these three measures, and thus provide evidence for them.

Accuracy as defined earlier is denoted by J and denotes the expected cost of discrepancies between actual and desired responses. An example of "desired" response would be the allocation of the most appropriate resource to handle a particular threat.

For timeliness there is an allotted time interval for the organizational response denoted by (Tmin, Tmax) where Tmax is an organizational threshold such that if the organization issues commands in response to the input after the threshold, there will not be enough time for implementation of an effective response. The expected time delay to produce a response is one measure that can be used to assess the timeliness of an organization's response. Another measure would be one that takes into account the variance of the response time. A useful measure is the probability that the organizational response lies in the desired time interval (Tmin, Tmax).

$$T = \text{Prob}(\text{Tmin} < \text{Td} < \text{Tmax})$$

To compute this probability density function (pdf) of the response time associated with particular decision strategies, all the information flow paths and the groups of concurrently active information flow paths in an organization must be identified. An algorithm for identifying the simple paths and the concurrently active paths has been developed by Jin, Levis, and Remy (1986). This provides for the computation of the pdf of the total delay of the organization. The details of that calculation appear in Andreadakis and Levis (1987) and Jin, Levis, and Remy (1986).

In the representation so far discussed, the human decision maker would be represented as part of a Petri Net. For a particular task, he might employ several processing algorithms as a function of varying inputs. This would be represented by several transitions and a decision switch appropriately linked. However, as earlier indicated, representing human processing algorithms and decisions to employ particular processing algorithms in such a manner greatly simplifies the complex human information-processing and decision-making processes. One way to introduce one of the constraints of human decision makers is to acknowledge that humans have limited information-processing capacities and requiring the human to process information at a rate higher than one which causes stress and performance decrements. This concept has been discussed repeatedly in research literature over the years (see for example, G. Miller, 1955; Tversky and Kahneman, 1974). The workload of decision makers is defined as the mental effort expended in order to perform their assigned tasks.

The workload depends on several factors including the uncertainty of the organization's task, the interactions among decision makers (organizational structure), the algorithms used to represent the various processing functions such as situation assessment or information fusion, and finally the internal decision strategies of each individual decision maker which involve probability of algorithms used. The idea that the rationality of the human decision maker is bounded (March, 1978) and that for each individual there is assumed to be a maximum information-processing rate is a basis for the concept that, in an organization, workload cannot exceed a threshold for each decision maker. It is therefore necessary to know what such a threshold is for different decision makers and different tasks and what happens to the excess workload that cannot be handled. This has been experimentally investigated by Louvet, Casey, and Levis (1988).

Using the Petri Net representation of organizations, the definitions and algorithms discussed provide for the mapping of organizational decision strategies into a locus of performance-workload space characterized by the three attributes: accuracy - timeliness - workload (J, T, G). Different organizational designs can be evaluated and compared on the basis of their corresponding loci. This was done in the experiment conducted by Andreadakis and Levis which simulated and compared parallel and hierarchical organizations in an air defense task involving three decision makers in each organization.

The accuracy (J) and workloads (G) of the three decision makers were computed for all decision strategies using algorithms implemented in software, to simulate the decision-making process and the protocols in each organization. The probability density functions (pdfs) of the response times of the two organizational structures were calculated and compared.

The authors concluded from their simulation that it is not possible to determine an optimal organizational structure unless one takes into account the parameters that characterize the task to be performed, e.g., rate of incoming information to be processed.

The Petri Net work described here demonstrates one attack on the important issue in C^3 evaluation, that of the human decision maker in C^3 organizations. Assessments of conditional probabilities of certain types of behaviors are difficult, especially where the situation dynamics are complex. Structural analytic techniques such as Petri Net provide promise for systematically modelling and experimenting on some of the elusive C^3 variables thus far discussed. Such measures as accuracy, timeliness, consistency, and the like can be translated to probabilities used in a Conditional Probability Logic approach to evaluating C^3.

It should be added that the work described is continuing through a DoD grant under the University Research Initiative to a George Mason University - Decision Science Consortium, Inc. (DSC) team investigating the impact of cognitive biases on command decisions and the potential for mitigating such impacts. The Petri Net approach is thus promising, but currently can be applied only to small C^3 organizations. Expansion to large-scale simulations and complex decision structures is still down the road.

A-5.0 SURVEY OF AVAILABLE TOOLS

This section addresses two topics. One involves available software packages for decision analysis, utility assessment, and other related topics discussed in this appendix. The second discusses some applications to military worth assessment not touched on elsewhere in the report. The survey of tools is only a cursory list of tools which might be investigated in a future assessment.

A-5.1 TOOLS FOR IMPLEMENTATION

The Decision Analytic paradigm has been discussed in detail in Sections A-3.1 to A-3.4. Decision Analysis software is available for decision trees, influence diagrams, and multi-attribute utility assessment (MAUT). Some of the products (e.g., ARBORIST - Texas Instruments; RISK - Palisade Corporation; and Decision 1-2-TREE - Fast Decision Systems, Inc.) are not very well-known. Others are available in commercial or prototype form and have been used enough for discussion. For example, Call and Miller (1990) provide a limited comparison of INDIA (Influence Diagrams - Decision Focus, Inc., Palo Alto, CA - beta test version), SUPERTREE (Decision Trees - Strategic Decisions Group, Menlo Park, CA), and DPL (Decision Trees and Influence Diagrams - Applied Decision Analysis Decision Systems, Menlo Park, CA). DAVID (Influence Diagrams, Ross Shachter, Stanford University) is a Macintosh based system. Call and Miller note that 1-2 TREE has a major flaw in that it cannot do Bayesian operations automatically. The RISK software cannot handle decision problems with multiple embedded decisions. Thus, Call and Miller limit their comparison to INDIA, DPL, and SUPERTREE.

Some of the typical issues that arise in the use of such software have to do with problem complexity. For example, for INDIA, the problem size that can be handled depends on the number of variables and the complexity of variable interrelations. For even small problems where most variables affected the value function, the PC based problem required the use of a hard disk. Value lotteries had to be specified as expected value nodes to reduce the problem burden. SUPERTREE is limited in the size of the problem it can handle with 8000 tree endpoints as an upper limit. This would limit the problem size to one equivalent to eight chance nodes each having three states where each state affects the value function. DPL is less limited than INDIA or SUPERTREE, and the reader is referred to the Call and Miller review (noting that they are members of the ADA staff).

HIVIEW has been developed specifically for Multi-Attribute Utility Theory (MAUT) and is available from the Decision Analysis Unit of London School of Economics (LSE). The software is user-friendly and provides for displays of tradeoffs among hierarchy attributes. It can be used for evaluating up to eleven options using an additive utility formulation. The number of attributes depends ultimately on the degree of decomposition of each hierarchy attribute. Also available from LSE is software known as EQUITY for cost-benefit types of analyses for resource allocation problems where the benefit measure can be multi-attributed (up to ten attributes).

Software also exists for implementing the Analytical Hierarchy Process (AHP). The software, Expert Choice, is PC based and is useful for obtaining weights in a hierarchy. It implements the AHP much as discussed in Section A-3.7.3.1.2 including the eigenvector approach to resolution of inconsistencies of preference.

A-5.2 MILITARY APPLICATIONS

This section will briefly discuss several applications to the evaluation of C^3 and will then explore the Joint Decision Support System in some depth.

A-5.2.1 Several C^3 Evaluation Paradigms

There have been several paradigms developed for the evaluation of C^3 effectiveness or C^3 requirements, many in support of the planning and budgeting process. Usually these paradigms are based upon the dichotomy of intelligence and operations functions. A major function during normal operations or combat is situation assessment which is handled by the intelligence function of the organization. The other major function, response, is divided into planning and implementation. The implementation is handled by the operations function and the planning is usually in a separate organization involving both functions because resource allocation in response to specific situations is required.

The situation assessment functions involve monitoring information, developing and evaluating explanations for information, and providing assessments to the operations function. The planning function usually develops an operations plan including resource allocations for specific scenarios that are hypothesized as possible or likely and updated as the situation develops. Finally, the operations function carries out the response to the situation. The C^3 part of all this is, of course, the entire process of assessing the situation, making the decisions, promulgating the commands, and ensuring continuity of communications and processing capability.

One paradigm for this process is the well-known SHOR paradigm usually attributed to Joe Wohl. This consists of:

STIMULUS - INCOMING DATA

HYPOTHESIS - PERCEPTION ALTERNATIVES

OPTION - RESPONSE ALTERNATIVES

RESPONSE - ACTION

The SHOR paradigm is obviously not unique in its representation of these functions, and it is more detailed in representing the interrelations of these functions than presented here. Other representations of the functions have labelled them variously as situation monitoring, data fusion, situation assessment, hypothesis selection, command interpretation, response selection, and response implementation.

Lawson has developed a decision process model that includes sensing the environment, processing information with respect to the status of friendly and enemy forces, comparing the assessed state with desired states, deciding on an action, and action implementation.

Still another process known as Headquarters Effectiveness Assessment Technique (HEAT) is an observation based method for evaluating the C^3 functions, in this case for a headquarters exercise. The theory of HEAT starts with overall C^3 functions such as monitor the environment, state estimation, situation assessment, option generation, option selection, plan generation and direction, and execution monitoring. The HEAT approach develops a detailed evaluation process involving data collection through observation of the headquarters exercise.

The HEAT approach then develops measures related to C^3 performance including queuing measures, soundness and coordination measures, quality measures, and characteristic measures with respect to implementation of decision making activities. Measures are developed for the functions described such as state estimation, situation assessment, planning, option generation, etc. Exercise observers are rigorously trained in data collection with respect to these measures to attempt to ensure consistency and inter-observer reliability. Data reduction forms are developed and functional scoring procedures developed that provide for analysis of overall C^3 performance.

HEAT is obviously one approach to evaluation of potential C^3 performance, through monitoring of simulated organizational performance. All the issues of validity, reliability, realism, etc., could be indicated but their existence is obvious. A relevant point is that this is indeed a rigorous process for simulating one kind of C^3 performance. How does it compare to other measures discussed such as simulations or requirement hierarchies? Other approaches, such as the Performance Assessment Report, a hierarchical roll-up approach, are discussed by Dockery (1988) in his review of case histories of C^3 evaluations.

The next section discusses a decision support system designed to support C^3 evaluators at the national command level, the Joint Decision Support System.

A-5.2.2 Joint Decision Support System

The Joint Decision Support System (JDSS) supports the global C^3 assessment, and it includes a set of analysis tools and a data base both dedicated to evaluating C^3 systems. The data base is written in ORACLE (as of March, 1990 - see Dockery, 1990) for personal computers using a 286 or 386 based chip architecture and running under MS-DOS. The JDSS uses WordPerfect for writing and Harvard Graphics for data displays.

JDSS is menu driven, and the user has three different response modes to the program. These are view only, modify, and analyze. There are seven different data bases including cost, programmatic information, lessons learned, and other information on C^3 systems and architectures. Most of the data is text material, but colors from red to green are used to represent subjective estimates with respect to sets of C^3 systems capabilities. JDSS serves the Joint Staff as well as action officers with the CINC (Commander in Chief) staffs worldwide. (For more information on the JDSS see Dockery, 1990) The remainder of this section provides a methodological critique of the JDSS.

The JDSS incorporates a method that its implementers consider to be ISMAUT. Neither the JDSS User's Manual nor its Programmer's Manual explains the axioms or foundations of JDSS nor do they provide references to other works that do. Therefore, we can only infer the foundations of JDSS from its descriptions. JDSS bears a surface resemblance to MAUT and ISMAUT, but it is without MAUT's foundations. In place of MAUT's well-developed formalisms, JDSS utilizes a combination of formalisms and procedures. JDSS contains not a "theory" so much as a combination of arbitrary procedures. Differences can be explored by examining and contrasting the methods used for evaluation and for roll-up in JDSS, MAUT, and ISMAUT. We will examine the characterization in the context of the hierarchy described in JDSS and its possible implementation with MAUT or ISMAUT.

A-5.2.2.1 Evaluation. A MAUT-based evaluation would be based on an overall value of programs as calculated in a MAUT analysis as explained above. Differences among JDSS prioritization, MAUT prioritization, and ISMAUT prioritization are summarized in Table 1, which utilizes, to the extent possible, the terminology of JDSS.

TABLE A-1. SUMMARY CHARACTERIZATION OF EVALUATION ALGORITHMS

FEATURE	JDSS	MAUT	ISMAUT
STRUCTURE	HIERARCHY	HIERARCHY	HIERARCHY
ALTERNATIVES	PROGRAMS	PROGRAMS	PROGRAMS
IMPORTANCE OF ELEMENTS	UNNORMALIZED VALUES ASSIGNED TO DEPENDENCIES	NORMALIZED VALUES ASSESSED ON RATIO SCALE	RANGES OF VALUES OR ORDERING RELATIONSHIPS
BOTTOM-LEVEL EVALUATIONS	VALUES ASSIGNED TO ASSESSMENTS AND NUMBER OF INSTANCES COUNTED FOR PROGRAMS	SCORES ASSESSED FOR CONTRIBUTIONS OF PROGRAMS TO CAPABILITY OBJECTIVE	RANGES OF SCORES ASSESSED FOR PROGRAMS OR DIRECT ORDERING RELATIONSHIPS ON PROGRAMS
OVERALL AND INTERMEDIATE VALUE	WEIGHTED SUM	WEIGHTED AVERAGE (ADDITIVE FORM)	WEIGHTED AVERAGE
EVALUATION (PRIORITY IN JDSS)	VALUE	VALUE	RANGE OF VALUES
OUTPUT	(COMPLETE PRIORITY ORDERING)	(COMPLETE PRIORITY ORDERING)	PARTIAL ORDERING (IF FEASIBLE)

All methods could use an identical hierarchy containing (from top to bottom): Global, CINC, Warfighting Environment, Major Mission Area, Mission Element, Level of Capability, Functional Task, and Capability Objective. With JDSS, the importance of sub-elements at each node in the hierarchy is assessed as essential, high dependence, moderate dependence, or contributing. These categories are assigned numeric values by the user or by default. With MAUT, weights are assessed on a ratio scale (e.g., an assessed value of 2 is judged to be twice as important as an assessed value of 1) and are then normalized at each level in the hierarchy. With ISMAUT, weights can be assigned as values on a ratio scale, intervals of values, or ordinal relationships. With JDSS, bottom-level evaluations are assigned to programs by: assigning numerical values to input assessment colors for capability objectives before and after SYDP, calculating the value difference, and summing across all capability objectives whose SYDP mentions the program. With MAUT, scores are assessed on a common ratio scale (typically based on a 0-to-100 point scale) of the contribution of the program to the improvement in each capability objective. With ISMAUT, scores can be assigned as values or ranges on a ratio scale or as ordinal relationships. Overall value or intermediate values (i.e., considering only certain branches of the hierarchy) are calculated as weighted sums by JDSS, with weights equal to the importance numbers. MAUT calculates overall and intermediate values as weighted averages by using the normalized weights and scores (for an additive structure). ISMAUT determines ranges of permissible values or relationships based on an underlying weighted average calculation. Evaluation in the cases of JDSS and MAUT is based on a program's calculated value; evaluation with ISMAUT is a range of values permitted by the input ranges and relationships on scores and weights. The output of JDSS and MAUT is a complete priority ordering of alternatives; ISMAUT provides information on the feasible partial orderings of alternatives.

There are four major shortcomings of JDSS with respect to evaluation of parameters:

1. A program's priority is driven by its number of occurrences rather than its contributions. A program that contributes very little to a lot of capability objectives will receive a higher priority than a program that contributes a lot to fewer capability objectives. This provides a vulnerability to the system. A program's advocate need

only argue that his program adds something, however small, to many capability objectives to improve its priority.

2. Assessed dependence is not a measure of relative importance. Since the numbers assigned to dependence categories are not normalized prior to the value calculations, they are not relative numbers. This flaw presents another vulnerability. In this case, a program's advocate needs to argue for high values of dependence for the branches leading to his program. This does not have to be at the expense of any other rating. Making everything "essential" in a portion of the hierarchy will inappropriately inflate everything below, relative to other parts of the structure. For example, if two major mission areas were both judged "highly dependent" for a warfighting environment but all of the mission elements within the first major mission area were assigned 'essential' while the second major mission area's elements were given a range of dependence assessments, then the first major mission area would be more important to the prioritization. Also, because weights are not normalized, an element that is subdivided into more sub-elements will be more important to the prioritization.

3. The simplified input structure encourages "garbage in," and it is impossible to improve the quality of poor input by manipulation and calculation alone. Several parts of JDSS's input are particularly susceptible to problems, including:

   a. There is no basis for JDSS (or users) to assign numerical weights to "dependence" assessments. The default procedure of assigning a value of 1 to 4 to the categories and then treating these numbers as if they were assessed as ratio-scaled judgments ("a weight of 2 will give all associated programs a resulting `value' twice that of a weight of 1," page 118 of the *JDSS User's Manual*), is just plain wrong. An extensive psychological literature in MAUT indicates conclusively that ratio judgments require care and that people do not normally associate the same value to verbal expressions. Further, JDSS provides no control over the numerical values assigned to the categories. For example, page 117 of the *JDSS User's Manual* shows a user how to change the numerical weight assigned to "moderate dependence" from the default value of 2 to a new value of 3, the same as "high dependence." This is done without changing the dependence assessment, so the user could promote his programs by changing weights and not having them even show up as a changed dependence.

   b. There is no basis for JDSS to assign numerical scores to assessment colors. This problem is similar to a. but worse because the *User's Manual* does not even state the values that are assigned to the colors. In addition, the convention of assigning a value of .5 (see p. 112 of the *JDSS User's Manual*) to cases where the SYDP will not result in a change in capability assessment is wrong. No improvement in capability should get no credit.

4. Priority is insensitive to cost. Priority is determined by the value of the program without regard to its cost. Cost is considered in the "Packages" option in JDSS, and it may distress a user to see that higher-priority items are omitted from packages. (This may just be a problem of semantics.)

ISMAUT does not suffer from JDSS's problems, but has a different set of weaknesses. The major practical weakness of ISMAUT, contrasted with MAUT, is the sheer volume of information that is required. Allowing users to specify ranges and relationships for input parameters may appear to reduce the elicitation burden but may actually create a greater burden. A potential problem could occur by encouraging respondents to think about "imprecise" specifications of parameter values; ISMAUT may also encourage respondents to provide more imprecise assessments. This may lead to respondents providing many more assessments than they would with MAUT. Although such responses would presumably be easier to make, their sheer volume, especially in a structure as complex as that used in JDSS, may be overwhelming. Additionally, the computational burden of keeping track of all of the imprecise relationships, especially in a large hierarchy, can slow the analysis dramatically.

The main theoretical problem with ISMAUT is that the input does not guarantee a complete ordering of alternatives. In the extreme case, a region may be infeasible and many of the specified relationships may need to be reconsidered and changed in order to produce feasibility. In a somewhat less extreme case, ISMAUT may provide no ordering of alternatives, inputs could be so imprecise that any ordering is permissible. In the usual case, ISMAUT will provide only a partial ordering. This may prove to be an unacceptable result if a lot of effort were required to provide input. ISMAUT might be improved to prompt users with an identification of places that need to be modified (e.g., by more precise specification of input) in order to provide a complete order of alternatives.

MAUT's major shortcoming is that it requires an exact specification of its parameters. It is often difficult for someone unfamiliar with MAUT to provide the inputs needed. A facilitator can help by posing questions in several different ways, by pointing out and seeking resolution of inconsistencies, by offering encouragement, and by iteratively displaying the results and implications of various assessments. An analyst can also keep the modeling processing moving when difficulty is encountered in the specification of a parameter. The analyst may start by using some of ISMAUT's techniques, but then suggest a specific value consistent with that information. Ranges, after attempting to reduce them, may be noted and used in subsequent sensitivity analyses.

It may even be possible to simplify some of the steps of MAUT (or ISMAUT) in the direction of JDSS without losing the foundations of MAUT. The following are some possibilities.

1. *Streamlined input of weights.* One can retain the method of indicating the relative importance of branches at each node by the assignment of categories, if this mode of input is desired by users. One could possibly even retain the categories of "essential" through "contributing." However, one would need to conduct research to assess from a sample of intended users the strength of importance that they attach to these phrases (e.g., is "essential" four times as important as "contributing"?). If agreement is found in these interpretations, one could incorporate a weight scale that reflects the agreed strengths of preference. If not, investigations are needed as to whether a scale can be improved for the categories. (We expect that an investigation would result in changes in either categories or numerical values or both.) Giving primacy to these ratio judgments rather than the terms themselves improves the accuracy of assessments. If categories are to be used for weights, then the user should not be allowed to change weights without changing categories. (This may require the addition of intermediate categories as is currently used with the colors.)

2. *Streamlined input of scores.* One could retain the procedures of assessing colors for assessment of capability objectives, if this is desired by the users. However, one would need to go through similar steps to determine scales. This numerical scale should also be made highly visible to the user. The assessment of scores involves two parts: an assessment of the capability desired (objective) and an assessment of the contribution of programs to the objective. The assessment of capability objective could use a 100-point scale with the following interpretations (this is just an illustration; the actual translation of the scale would need to be determined): 0-20 = green, 21-40 = yellow/green, 41-60 = yellow, 61-80 =

red/yellow, 81-100 = red. This range of scales could also be used in the roll-up algorithms as described below. For purposes of the prioritization, mid-point values in the ranges could be assigned to the assessed colors. Thus, red would be assigned 90, red/yellow = 70, yellow = 50, yellow/green = 30, and green = 10. The numbers in the example are appropriate only if the transition from one color to the next is of equal value (e.g., if going from red/yellow to yellow is as valuable as going from yellow to yellow/green). An assessment of each program's contribution to the SYDP's improvement in capability objectives is also required. This is a new assessment, but it is necessary. This might simply be an assessment of high, medium, or low contribution, possibly with the option to specify the same contribution for a program to all capability improvements where it appears. In this case, a ratio scale could also be introduced (with a verified or required interpretation). A possible result might be that high is regarded as three times the contribution of low and that medium is twice that of low. If even this is too burdensome, each program's contribution could be counted as equal, but a threshold should be imposed for including a program. In either case, the contributions of programs to each capability objective should be normalized.

A-5.2.2.2 Roll-Up. Roll-up in MAUT is performed as the successive calculation of weighted average values as described above. Differences between JDSS roll-up, MAUT roll-up, and ISMAUT roll-up are summarized in Table A-2.

TABLE A-2. SUMMARY CHARACTERISTICS OF ROLL-UP ALGORITHMS

FEATURE	JDSS	MAUT	ISMAUT
STRUCTURE	HIERARCHY	HIERARCHY	HIERARCHY
IMPORTANCE	DEPENDENCIES	NORMALIZED WEIGHTS (FOR ADDITIVE STRUCTURE)	RANGES ON NORMALIZED WEIGHTS OR ORDINAL RELATIONSHIPS
BOTTOM-LEVEL EVALUATIONS	ASSESSMENTS	SCALES	RANGES ON SCALES OR ORDINAL RELATIONSHIPS
COMBINATION RULE	AD HOC RULES	WEIGHTED AVERAGE (FOR ADDITIVE STRUCTURE	WEIGHTED AVERAGE
INTERMEDIATE ROLL-UPS	PROVIDED	PROVIDED	NOT CLEAR

All methods can use identical hierarchies. JDSS uses dependence categories (essential, high dependence, moderate dependence, contributing) to characterize the importance of sub-elements at each level in the hierarchy; MAUT uses normalized weights; ISMAUT uses ranges on normalized weights or ordinal relationships. JDSS uses capabilities assessments (red, yellow, green) to characterize bottom-level evaluations; MAUT uses value or utility scales; ISMAUT uses ranges on scales or ordinal relationships among attributes. JDSS uses an ad hoc set of rules to combine values of branches into a node value; MAUT uses weighted averages. Both JDSS and MAUT provide roll-up values at intermediate levels of the hierarchy. It is not clear what ISMAUT does; it has enough information to provide ranges of roll-up values or dominance information based on the partial (rolled-up) list of attributes.

There are four major shortcomings with JDSS with regard to roll-up of evaluation parameters:

1. *There is no basis for the set of combination rules.* Specific rules are given on pages 130 and 131 of the *JDSS User's Manual* and on pages 182-183 of the *JDSS Programmer's Manual.* Combination rules depend on the highest dependence level of branches at each node and the range of assessment colors. A tentative rolled-up value is initially set equal to the worst or median assessment of a sub-set of branches. This tentative assignment might then be "pushed-down" or it might be degraded one shade. The source of the rules is not specified. Furthermore, the author of the *JDSS User's Manual* (who is not named) appears to disclaim responsibility for the rules on page 129 by stating, "the rules listed on the following pages are intended to promote discussion throughout the community." No disclaimer is on the computer screens, however. There is no axiomatic basis for the rules, there is no source given for the rules, the author disclaims the rules, yet they are presented without warning in JDSS. The complete effects of this system of rules have clearly not been explored in either the *JDSS User's Manual* or the *JDSS Programmer's Manual.*

2. *Combination rules are conservatively biased.* At each step in the hierarchy with an "essential" or "highly dependent" component, the rolled-up value is initially assigned to the worst of these components. This initial assignment may then be "pushed-down," and the new assigned value could be further degraded by one shade. In the extreme, a single red value in a string of essential branches could cause the whole assessment to be red. This "weakest link" line of reasoning seems extreme, especially without additional justification.

3. *Roll-up is not linked to evaluation.* JDSS's evaluation algorithm is similar to an additive MAUT's algorithm (but with the problems addressed above). However, JDSS's conservative roll-up rules are far from additive (additive rules, for one thing, are compensatory). If the extreme, "weak link" sorts of combinations are really justified, they should drive the evaluation as well, elevating programs that fix the weak links rather than those that contribute to nodes that may not even influence the roll-up at all. In addition, if something other than an approximately additive combination rule is appropriate, then the procedure of top-down assessments of "dependencies" is incorrect.

4. *The rule set is cumbersome.* This adds to the processing time compared with an analytical form of the MAUT analysis, but may be no worse than ISMAUT.

The descriptions of ISMAUT do not provide such information on roll-up. However, since it is MAUT-based, we suggest that roll-up could be supported in the same manner as MAUT provides roll-up. This provides a logical tie between evaluations and roll-up. ISMAUT would retain the computational complexities in roll-up that it has in evaluation.

An additive MAUT formulation overcomes most of the problems of JDSS. Its basis can be made explicit and justified, its combination rules are not biased, roll-up is linked to evaluation, and calculation is easy. An additive formulation is often an adequate approximation to more complicated forms, especially when the hierarchy is extensive (as is the case with JDSS). However, the additive algorithm is compensatory, i.e., deficiencies in one area can be compensated for by performance in other areas. The adequacy of the formulation needs to be justified. If the additive formulation is inadequate, MAUT offers other combination rules. The axiomatic basis has been developed for many combination rules including: multiplicative, utility-independent, generalized multiplicative, bilaterally independent, and joint interpolation-independent forms, among others (including the general formulation). These other forms are more complicated to deal with and will slow calculations.

If none of the analytical forms of combination rules is appropriate, then a rule-based algorithm could be used in conjunction with MAUT. However, the development of such an algorithm should be performed in a defensible manner, unlike the algorithm in the present JDSS. The first step in developing a defensible set of rules is to identify experts who are appropriate for making judgments about how assessments should combine. Experts might be different at different levels in the hierarchy. For example, each CINC (or his staff) may be most appropriate for judging the contribution of major mission areas to warfighting environments and the contribution of mission elements to mission areas. Technical specialists may be better for judging the contribution of functional tasks to levels of capability. Next, combination rules must be elicited from the experts. This may require a number of individual elicitation methods such as: knowledge elicitation, paired elicitation methods such as consensus groups and nominal groups, to elicit the most correct and defensible information. The sources of these judgments should be documented. Finally, the rules may be refined by iterating with the experts. Iterations should: a) point out vulnerabilities and biases in the rules that can be determined by analysis; b) show the results of applying the rules to actual previous assessments; and c) show the implications of the rules in hypothetical assessments.

A-5.2.2.3 Packages. Both the ISMAUT+ computer program (Stewart et al.) and JDSS allow for users to specify additional requirements on choices of alternatives to recommend "packages" or groups of alternatives. Some requirements could be constraints including: a budget on the total cost of alternatives, a requirement to include an alternative, a requirement to include only one of a list of alternatives (exclusive OR), a requirement to include at least one of a list of alternatives (inclusive OR), or a requirement that an alternative be excluded from the package. ISMAUT+ also allows a requirement to include all of a list of alternatives (AND), or a requirement that the package provide a minimum level of contribution to an attribute.

The ISMAUT+ program also allows a user to specify positive or negative "synergy scores" on attributes between pairs of alternatives. This feature accounts for complementary or supplementary performance of alternatives. (The assignment of benefit scores in this part of ISMAUT+ appears to require an assignment of a specific set of values rather than a range or ordinal relationship).

JDSS claims to use the ISMAUT+ algorithm to create a package (see *JDSS User's Manual*, p. 146). However, there appear to be some discrepancies in the written descriptions of the algorithm. JDSS claims to start "by creating an arbitrary package of programs that satisfies all program selection constraints," but not the budget constraint. If this exceeds the budget, then another package is analyzed until one is found that meets the budget. JDSS then adds programs in order of decreasing value to cost ratio until the budget is met. The process is then repeated "to see if another package can be created with a higher overall value." Then, "after all of the possible packages have been considered, the JDSS selects the package with the highest value," presumably using its evaluation method (which would leave packages disconnected from roll-up in the same way that evaluations are disconnected). It is unclear why JDSS iterates in this manner if it really evaluates a complete enumeration of packages ("all of the possible packages"). This is an extremely cumbersome and time-consuming approach, which, if taken at its word, may even be infeasible in some cases (because there are so many possible packages). In any case, this is a problem (known as the "knapsack problem") that has been studied for a long time by the operations research community. Integer programming algorithms (such as branch and bound) have been developed to solve these problems efficiently, and it is a mystery why these methods were not included in JDSS. (However, even these algorithms are quite cumbersome and time-consuming.)

ISMAUT+ (Stewart et al., pp. 23-25) considers all constraints simultaneously and generates all feasible packages (again, complete enumeration). It then calculates values for the packages against all attributes considering the synergies. This information can then be used by ISMAUT for its type of evaluation, i.e., identification of dominance relationships and calculation of ranges of evaluations for packages. Stewart et al. recognize that "the ability of the ISMAUT methodology to handle the resulting enormous numbers of alternatives is quite limited." This is clearly the case for a structure as complex as the one in JDSS.

MAUT is not especially adapted to solve constrained optimization or budgeting problems, nor is it addressed at evaluating packages of alternatives. However, with some modification, MAUT can be used to address the type of packaging problems that are addressed by ISMAUT and JDSS. The following procedure can modify an additive MAUT procedure to handle a cost constraint (budget), benefit synergies, cost synergies, and OR, AND, inclusion, and exclusion constraints.

Traditional additive MAU analysis evaluates programs on the basis of their weighted-average performance against attributes. This is a satisfactory method for determining the "best" program and its benefit relative to other programs when a single program is to be chosen. This procedure falls short, however, when several programs are to be chosen. An analytical method to assist in the choice of several programs (a package) must address other aspects, especially budgets and interactions among programs. The budget can be addressed by ranking the programs on the basis of their benefit-to-cost ratios. This provides a prioritization that maximizes the benefit achievable for any given budget. (In some cases, this simple procedure has approximation error, but the error is small when many programs (greater than about 25) are involved.) Ranking on the basis of benefit-to-cost ratio is fast, easy, and provides the highest valued package for many budget levels.

The method is complicated, however, when programs interact. In this case, the benefit from any given program may depend on which other programs are already selected. Thus, the benefit component of the benefit-to-cost ratio changes, and a prioritization based on static benefit estimates can be seriously in error. This problem can be addressed by developing a prioritization algorithm that reflects the first-order value and cost interactions between pairs of programs. This feature could allow the user to specify the interactions of pairs of programs on attribute scores and program costs. Value interactions may be complementary (the score of the pair of programs is greater than the sum of their individual scores) or supplementary (the score of the pair is less than the sum of individual scores) and may differ by individual attribute. This procedure is similar to the "synergy" assessment in ISMAUT+. Cost interactions may also be complementary or supplementary. This change is then accommodated in the prioritization algorithm by choosing programs based on their benefits and costs incremental to the set already specified. That is, the second priority would be determined from an analysis based on the benefit and cost of each program over and above that provided by the highest-priority program. The third priority would be chosen on the basis of the benefit and cost of each program over and above that provided by the two highest programs, and so forth. The incremental algorithm is slower because the calculation of benefit-to-cost ratio is repeated after each selection, but it is still considerably faster than generating the set of feasible packages.

Most packaging rules of the type accommodated in ISMAUT+ and JDSS are easily accommodated in the MAUT prioritization algorithm with little or no adverse effects.

1. *Inclusion*. The user may specify that any number of the programs must be purchased. In effect, these are given top priority regardless of their costs and benefits and so are moved to the top of the list automatically without calculation.

2. *Exclusive OR*. The user may specify that, of a list of programs, only one is to be prioritized. Once one of these programs is prioritized, the other will be put at the end of the list and eliminated from further calculation. This rule could be used in conjunction with "inclusion."

3. *Inclusive OR*. This function may be used *only* in conjunction with "inclusion." The user may indicate that one of list of programs must be included. The algorithm would then select the one offering the highest benefit to cost.

4. *AND*. The user may specify groups of programs that must be purchased together. The system would use the combined benefit and cost of the groups when determining priorities.

A specification of a minimum required level of performance on an attribute is more difficult. One possibility is to run the packaging algorithm without this constraint and check, after-the-fact, to see if it is satisfied. If it is not, then the lowest priority programs could be dropped from the list and replaced with programs that were excluded form the original list but which provide the highest contributions to the deficient attribute. The procedure contains some approximation error, but it is much faster than complete enumeration.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B

RESEARCH ISSUES
CONCERNING
DECISION MAKING UNDER UNCERTAINTY

B-1.0 RESEARCH ISSUES

A number of questions arose in the course of the research which suggested further research was required.

(a) Worth Independence vs. Conditional Independence: For a particular context, we may be able to identify a causally conditionally independent set of variables (the objectives), but what are the requirements to be consistent with the objectives at the next level? (Causal conditional dependence is equivalent to saying there is an underlying functional relationship from one (antecedent input) to another (consequent output). See the discussion of causality and conditional dependence below.)

Conjecture: Worth is additively independent if and only if the objective criteria are (thought to be) causally conditionally independent. (See Keeney and Raiffa for independence requirements for additivity.) We already know (section 2.2.2 of the main report) that conditional independence of the outcome, given the alternatives, simplifies the calculation of the utility independent (multiplicative) form.

The issue becomes how to explicate and relate perceptions of worth at antecedent variables, such as "capabilities", compared to perceptions of worth at the level of the objectives (Mission Success Criteria) and how do the latter relate to the worth for a higher level of objectives? The discussion of proxy variables is another form of this question. Should variables be conditionally independent to be commensurate with respect to worth? Can we elicit worth on antecedents and consequents? Are they separable and do they obey equations of the conditional utility function (see equation (10) of the main report) or conditions for additive independence?

For example, the number of capital ships and support ships on both sides are conditionally independent of each other (for a point in time). An earlier (antecedent) value of those variables may be antecedent to their later outcome, but not simultaneously.

An example of simultaneous causal dependence is that the number, R, of remaining (surviving) units depends on the sum over the survived/destroyed state (variable) of each unit at a point in time for all combinations of those states that add to "r" for all values of R. Do the antecedent surviving/destroyed states have an intrinsic worth that is separable from the worth on the number of remaining units? From an unemotional point of view, perhaps not, but if a particular unit has special significance, then, of course, it may have special worth. Should it be included in the worth of the total remaining units other than the special one?

Another example of simultaneous dependence is the utility sum of individual utilities, provided they are simultaneously dependent on the (attribute) variables for which they are applied, and provided that they are additive independent utilities.

(b) Sensitivity Analysis: The sensitivity of the result to changes in performance of one capability depends on the set point of the other capabilities, i.e., the partial derivative of the functional* of the outcome with respect to the capability in question depends on the effectiveness** function(s) attributed to the other capabilities.

*The functional may be the instance value of the variable(s), the probability function on the outcome variable(s), the probability function on the worth of the outcome, or the expected utility .

**Effectiveness functions are conditional probability functions of the "capability" variable, e.g., detection, conditioned on the variables that effect that performance, such as, range.

(c) Utility Variance: The MAUT approach is predicated on expected utility. How does the distribution of utility, and, in particular, the utility variance come into play?

Conjecture: Expected utility is not sufficient to predict the behavior of a decision maker, because the decision maker does not actually face a large number of repetitions to average out his wins and losses. The decision maker faces a sample of one. What is the error function for a sample of one? How does the spread (variance, entropy) of the probability of worth effect the decision?

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1 AGENCY USE ONLY *(Leave blank)*	2 REPORT DATE November 1991	3 REPORT TYPE AND DATES COVERED Final: Oct 90 – Sep 91

4 TITLE AND SUBTITLE

MODELING CHOICE UNDER UNCERTAINTY IN MILITARY SYSTEMS ANALYSIS

5 FUNDING NUMBERS

Program Element: 605853N
Project: CD87
Task: R 0905
Work Unit: 305 162
Agency Accession No.: DN305162
Contract No.: N66001-90-D-0192

6 AUTHOR(S)

P. E. Girard, Science Applications International Corp.; M. F. O'Connor, J. W. Ulvila, Decision Science Consortium, Inc.

7 PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Ocean Systems Center
San Diego, CA 92152-5000

8 PERFORMING ORGANIZATION REPORT NUMBER

NOSC TD 2216

9 SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Chief of Naval Operations
Deputy Chief for Naval Warfare (CNO-OP-76)
Washington, DC 20350-2000

10 SPONSORING/MONITORING AGENCY REPORT NUMBER

11 SUPPLEMENTARY NOTES

12a DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b DISTRIBUTION CODE

13 ABSTRACT *(Maximum 200 words)*

This report presents the rationale for a method of making effectiveness and worth judgments of warfare systems, explicitly addressing command and control, based on Multi-Attribute Utility Theory. The elements of the problem are identified and the analysis places into perspective the aspects of objectives, system capabilities, choice alternatives, and their relative worth. Research issues are also identified.

14 SUBJECT TERMS

command and control, Multi-Attribute Utility Theory, military systems analysis, uncertainty, modeling, choice, warfare systems effectiveness and worth

15 NUMBER OF PAGES

94

16 PRICE CODE

17 SECURITY CLASSIFICATION OF REPORT	18 SECURITY CLASSIFICATION OF THIS PAGE	19 SECURITY CLASSIFICATION OF ABSTRACT	20 LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAME AS REPORT